Functional characterization of allelic variants of pulmonary adenoma susceptibility 1 (Pas1) genes for their cancer modifier activity

Thesis

UNRESTRICTED

Federica Galbiati

Degree in Biological Science

FUNCTIONAL CHARACTERIZATION OF ALLELIC VARIANTS OF PULMONARY ADENOMA SUSCEPTIBILITY 1 (*Pas1*) GENES FOR THEIR CANCER MODIFIER ACTIVITY

Thesis Presented To

The Open University of London for the Degree of

Doctor of Philosophy

Discipline: Life Sciences

February 23rd 2007

Sponsoring establishment:

Istituto Nazionale Tumori, Milan (Italy)

DATE OF SUBMISSION 05 DECEMBER 2006
DATE OF AWARD 31 OCTOBER 2007

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

INDEX

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The pulmonary adenoma susceptibility 1 (*Pas1*) locus affects inherited predisposition and resistance to chemically induced lung tumorigenesis in mice. A/J and C57BL6/J mouse strains carry the susceptibility and resistance allele, respectively. In the present study we identified and genotyped 64 polymorphisms in the *Pas1* locus in 29 mouse inbred strains in order to reduce the region containing the *Pas1* locus, and we succeeded in delimiting the *Pas1* locus to a minimal region of 468kb containing six genes (*Bcat1*, *Lrmp, Casc1*, *Ghiso*, *Kras2* and *Lmn-rs1*). That region defined a core *Pas1* haplotype (A/J- and C57BL/6J-type) with 42 tightly linked markers, including intragenic polymorphisms in five genes (*Bcat1*, *Lrmp, Casc1*, *Ghiso* and *Kras2*) and amino-acid changes in three genes (*Lrmp*, *Casc1* and *Lmn-rs1*). mRNA expression study of these candidate genes revealed that, in (A/JxC57BL/6J)F1 mouse lung tumors, the *Lmna-rs1* gene was completely downregulated, whereas allele-specific downregulation of the C57BL/6J-derived allele was observed at the *Casc1* gene, suggesting the potential role of these genes in tumor suppression. These first results indicate a complex multigenic nature of the *Pas1* locus, and point to a functional role for both intronic and exonic polymorphisms of the six genes of the *Pas1* haplotype in lung tumor susceptibility.

Then we continued by testing functional activity of candidate genes in vitro. In vitro colony formation assay of human lung cancer cell lines A549 and NCI-H520 transfected with the allelic variants of the four genes revealed allele-specific modulations of colony numbers by *Lmna-rs1* and *Casc1*, but not by *Lrmp* and *Ghiso.* We found that inhibition of clonogenicity by allelic forms of *Pas1* candidate genes was not mediated by induction of apoptosis.

These findings provide evidence that allelic variants of mouse *Pas1* candidate genes differentially modulate growth of human cancer cells.

Chapter 1 INTRODUCTION

1.1 CANCER GENETICS: BASIC KNOWLEDGE

Cancers are derived from single somatic cells and their progeny in which regulatory mechanisms for proliferation have been disrupted (Ponder, 2001). Malignant cells have several qualities that distinguish them from their normal counterparts. They are immortal, often grow more rapidly than normal cells of the same origin, and fail to exhibit normal cell-cell interactions (Bale and Li, 1997) Tumorigenesis in humans is a multi step process. Steps or "hits" are mutations or other alterations of genes involved in regulation of cell growth and cell interactions. Mutations and alterations may occur by random chance during DNA synthesis and cell replication, after an exposure to environmental carcinogens, or they may be inherited as germline mutations.

Hanahan and Weinberg in 2000 suggested that, in perhaps all types of human tumors, cancer cell genotype is a manifestation of six essential alterations in cell physiology that collectively dictate malignant growth (Fig.1):

1. Self-Sufficiency in growth signals. Normal cells require mitogenic growth signals (GS) before they can move from a quiescent state into an active proliferative state. Many of the oncogenes act by mimicking normal growth signaling.
2. Insensitivity to antigrowth signals. Multiple antiproliferative signals operate to maintain cellular quiescence and tissue homeostasis in a normal tissue. Cancer cells must evade these antiproliferative signals if they are to prosper.
3. Evading apoptosis. Programmed cell death - apoptosis- together with the rate of cell proliferation determines the ability of tumor cell population to expand in number. Cancer cells can acquire resistance to apoptosis.
4. Limitless replicative potential. Perhaps all types of mammalian cells carry an intrinsic, cell-autonomous program that limits their multiplication. This program must be disrupted in order

for a clone of cells to expand to a size that constitutes a macroscopic life-threatening tumor.

5. Sustained angiogenesis. The oxygen and nutrients supplied by the vasculature are crucial for cell function and survival, obligating virtually all cells in a tissue to reside within 100 μm of a capillary blood vessel. The ability to induce and sustain angiogenesis seems to be acquired in a discrete step (or steps) during tumor development, via an "angiogenic switch" from vascular quiescence.
6. Tissue invasion and metastasis. Primary tumor masses spawn pioneer cells that move out, invade adjacent tissues, and thence travel to distant sites where they may succeed in founding new colonies.

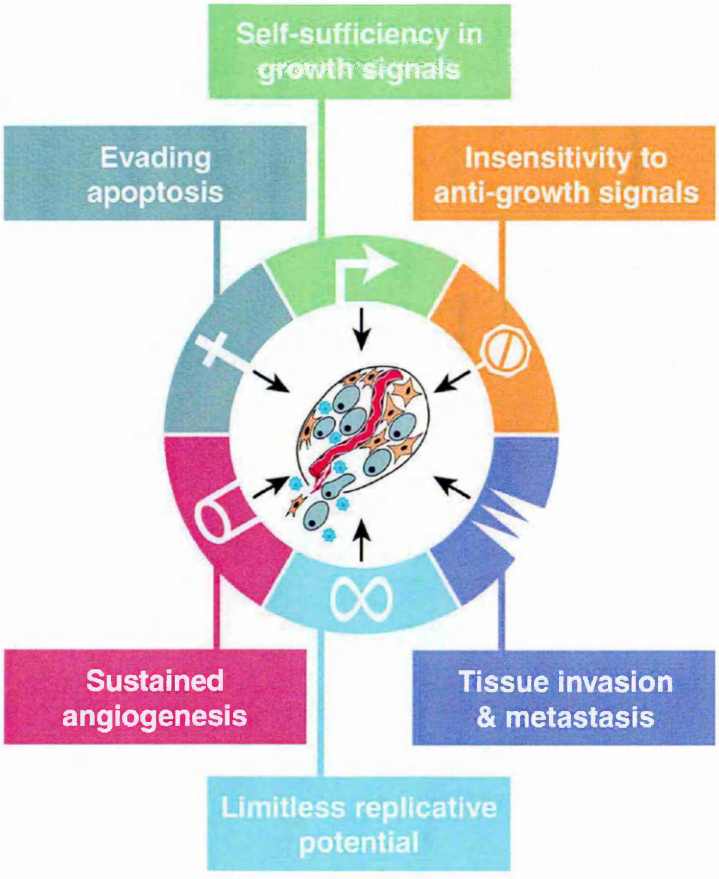

Figure 1 Acquired capabilities of cancer. *From Hanahan and Weinberg, 2000.*

The order in which these capabilities are acquired seems likely to be quite variable in all cancer types and subtypes. A particular genetic lesion may confer several capabilities simultaneously, decreasing the number of distinct mutational steps required to complete tumorigenesis.

Otherwise, when a capability is acquired through the collaboration of two or more distinct genetic changes, the total number necessary for completion of tumor progression is increased.

Several factors can influence the evolution of a cancer.

These factors reflect genetic or epigenetic alterations, each conferring one or another type of growth advantage, that drive the progressive transformation of normal cells into highly malignant derivatives. These events are somatic events but in many of the inherited cancer syndromes one of the events is inherited.

1.1.1 Inherited and sporadic cancer

Some individuals are genetically predisposed to a particular type of cancer (Ponder, 1990). Most common hereditary cancers are sporadic and the genetic alterations responsible for their genesis take place over several cell generations; nevertheless, 5 to 10% of human tumors are hereditary, with a rapid development (Russo *et al.*, 2000). Family history is important to recognize clinically inherited predisposition. A family history of cancer, early age of onset and absence of known carcinogen exposure, suggest an underlying genetic disorder. In terms of family history, it is possible to distinguish three categories of inherited predisposition to cancer. (Table 1 in Ponder, 2001)

Table 1 Inherited predisposition to cancer

	Contribution to overall cancer incidence	Clinical features	Frequency of predisposing alleles	Effect on individual risk
Inherited cancer syndromes	1–2% at most	Rare/unusual cancers or combinations of cancers. Sometimes with associated developmental defects or non-neoplastic phenotype. Mendelian dominant inheritance	Rare (~1:1,000 or less)	Strong lifetime risks of cancer up to 50–80%
Familial cancers	Up to 10% depending on definition	Families with several cases of common cancers that fall into a recognized pattern of cancer types (for example, breast and ovary; colon–endometrium–ovary). Spectrum from families with multiple cases at young age (strongest evidence of predisposition) to two or three cases at older ages; many of the latter will be due to chance or to combinations of weaker genes. Generally show pattern consistent with dominant inheritance	Uncommon to common	Moderate to weak
Predisposition without evident family clustering	No precise figure possible. Distribution of risk within population may result in substantial fraction of cancer incidence within predisposed minority	Single cases of cancer at any site, some with one or two affected relatives. The distribution of these cases in the population is probably determined by the combined effects of multiple genetic and non-genetic risk factors	Multiple common alleles	Weak

Table 1 Inherited predisposition to cancer. *From Ponder, 2001.*

- Inherited cancer syndromes: the incidence is 1-2% at most. These syndromes consist in rare or unusual cancers or combinations of cancer. Sometimes these syndromes are associated with specific phenotypes that are attributable to the predisposition gene. A predisposing mutation is present in all somatic cells and in the germline, which enables the transmission of risk to the next generation.
- Familial cancers: the incidence is up to 10%. This category includes cancers that tend to cluster in families but, because of the lack of markers for the predisposed phenotype, there may be doubt about the presence of inherited predisposition in any given family.
- Inherited predisposition without evident family clustering: in this case, there is an inherited disposition with low penetrance so that familial clusters are rare. The distribution of these cases in the population is probably determined by the combined effects of multiple genetic and non-genetic risk factors.

In the last few years a number of genes, whose germ-line mutations cause a highly penetrant familial predisposition to cancer have been mapped on the human genome, and some have been cloned and

identified. In humans the frequency of alleles that confer an inherited predisposition to cancer is variable, being about 1:100,000 in the case of retinoblastoma (RBI), 1-2:10,000 in neurofibromatosis (NFI), 1:1,000 in familial breast cancer (*BRCA1*), and 1:1,000 in hereditary nonpolyposis colorectal cancer (HNPCC).

The greatest discoveries to date in the genetics of common inherited cancers relate to breast, ovarian, and colorectal cancer. In particular there have been studies of mutations in two genes for breast and ovarian cancer (*BRCA1* and *BRCA2*), the *APC* gene for familial adenomatous polyposis, and several mismatch repair genes, which correct errors in DNA that arise after replication, for hereditary non-polyposis colon cancer (Emery *et al.*, 2001).

All cancers are genetic in origin because they arise from mutations in a single somatic cell, but the genetic changes in sporadic cancer are confined to a particular tissue. In inherited cancers, a predisposing mutation is present in all somatic cells and in the germline, which enables the transmission of risk to the next generation. Mutations in *BRCA1*, *BRCA2*, *APC* and the genes related to hereditary non-polyposis colon cancer are inherited in an autosomal dominant fashion. However, *BRCA* and the genes associated with hereditary non-polyposis colon cancer (and a few *APC* mutations) show incomplete penetrance, and they predispose a carrier to develop specific cancers. Additional genetic, environmental, and hormonal factors determine the ultimate development of cancer in mutation carriers (Emery *et al.*, 2001).

The pattern of genetic alterations differs between cancer of different types, and of the same type. This genetic pattern is not random and it is influenced not only by genetic background, but also by environmental exposures.

Genetic variation that acts either within or outside the cancer cell may determine the outcome of interaction with exogenous carcinogens. An example is provided by the risk of cutaneous

melanoma which is increased by exposure to the sun in individuals with a fair skin or who have many naevi.

1.1.2 Polygenic inheritance of predisposition to cancer

Polygenic inheritance of predisposition to cancer is demonstrated in experimental animals for different tumor types. Genetic susceptibility to different types of cancer is determined by inheritance of multiple cancer predisposition and resistance alleles, whose chromosomal locations have been found by genetic linkage analysis.

Epidemiological evidence indicates that in many cases of sporadic cancers, first degree relatives of cancer patients show an excess risk of cancer when compared with the general population. This evidence may suggest the existence of a polygenic inheritance of predisposition to particular types of cancer (Dragani *et al.*, 1996).

Several types of genes are involved in carcinogenesis. Normal cell regulation results from balance between the function of growth-promoting genes and growth-suppressing genes. When the former type of gene is "activated" through amplification or mutation to a hyperfunctioning form, it exerts a dominant positive effect of cell growth. Such genes are known as "oncogenes". The latter type of gene is known as a "tumour suppressor". Loss of function of both homologous copies of a tumour suppressor gene has a powerful growth-promoting effect. Loss of just one copy typically has very little effect. Other genes that may be altered in cancer cells include those involved in DNA repair and in programmed cell death.

Tumour suppressor genes were identified through their role in familial predisposition to cancer. These genes are known to be critically important in growth control of both sporadic and hereditary

tumours. The majority of hereditary cancer predisposition is attributable to germline mutations in tumor suppressor genes.
Expression of protooncogenes in normal cells appears to be carefully regulated. Different oncogenes are turned on/off at different steps of the cell cycle, at different stages of embryologic development, and in different cell types. Protooncogenes function as messengers in the pathway by which external stimuli received at the surface lead to DNA synthesis, cell growth, and division (Bale and Li, 1997).

1.1.3 Cancer modifier genes

Variation in phenotypic expression is a feature of many dominantly inherited single-gene disorders. An example of this kind of disorder is Neurofibromatosis type 1 (NF1) that is notable for its variable expressions. Although the major diagnostic features of cafè-au-lait patches, cutaneous neurofibromas, and Lisch nodules are each present in more than 90% of cases of puberty, the number of lesions varies greatly. Furthermore, there are a number of characteristic features which only occur in a minority of cases such as severe learning difficulties, scoliosis, optic gliomas, epileptic seizures, and pseudoarthrosis. Some of this variation in phenotypic expression could reflect different mutations in the NF1 gene. Almost half of all NF1 cases are the results of new mutations and a number of different mutations have been reported. Variation in the NF1 mutation is not the only source of variation, because there are also considerable differences in expression between members of the same family. In 1979, Carey *et al.* demonstrated that three-quarters of families showed marked differences in severity of NF1 between individuals. The remainder of the phenotypic variation could be due to other modifying genes, environmental factors, or some combination of these. The term "modifying gene" is used to denote

any gene unlinked to the NF1 locus whose genotype correlates with the NF1 phenotype (Easton *et al.*, 1993).

Nathanson *et al.* in 2002 observed that the substantial variability in penetrance of breast cancer due to germline mutations in *BRCA1* has led to speculation that inheritance of *BRCA1* mutation alone may not completely describe the observed cancer phenotype. Considerable variation in the degree of breast cancer risk associated with a *BRCA1* mutation has been observed and this suggests that modifiers of *BRCA1* penetrance may exist.

Cancer modifier loci are defined as genetic loci involved in fine-tuning the tumor phenotype through allele specific effects on cancer development, e.g., stimulation or inhibition of tumour development and/or tumour progression. Cancer modifier loci may comprise a "null" allele, with no effects on tumour phenotypes, and a "susceptibility" or "resistance" allele, which affect tumour phenotype in a dominant, codominant or recessive way.

The study of mouse cancer modifier loci might serve as the basis for understanding the genetic and biochemical mechanisms of polygenic inheritance of cancer predisposition/resistance. Identification of homologous cancer modifier loci in humans might provide a step toward the development of diagnostic, preventive, and therapeutic strategies that target these loci (Dragani, 2003). Cancer modifier loci affecting specific stages of the tumorigenesis process have been mapped in the experimental models where the tumor-initiation and tumour-promotion/progression phenotypes are clearly defined. For example, in lung tumorigenesis, the *Pas1* and *Par1* loci affect both tumour multiplicity and size, whereas *Par2* and *Par4* act specifically on tumor multiplicity, and *Papg1* and *Sluc* affect only tumor size. (Gariboldi *et al.*1993, Manenti *et al.*, 1996 and Tripodis *et al.*, 2001 in Dragani, 2003) Cancer modifier loci may affect either single or multiple stages of carcinogenesis.

1.2 GENETIC DETERMINANTS OF POLYGENIC TRAIT

The major challenge for medical genetics for the next decade is to systematically identify genes responsible for the hereditary contributions to a range of disorders that are not inherited in a simple Mendelian fashion but that have a definite or probable genetic component. Complex diseases are thought to involve susceptibility genes at multiple loci, known or unknown environmental factors and, commonly, incomplete penetrance of susceptibility genes. This "complex" trait includes a large fraction of the common causes of disability and death in Western society such as atherosclerosis, hypertension, psychiatric disorders, Alzheimer disease, type I and type II diabetes, asthma, rheumatoid arthritis, and probably significant components of the general aging process. These disorders may be oligo- or polygenic, and environmental and stochastic developmental and immunologic processes may play a major role in the development of these diseases (Weissman, 1995).

The term "complex trait" refers to any phenotype that does not exhibit classic Mendelian recessive or dominant inheritance attributable to a single gene locus. In general, complexities arise when the simple correspondence between genotype and phenotype breaks down, either because the same genotype can result in a different phenotype or different genotypes can result in the same phenotype (Lander and Schork, 1994).

There are some problems in the study of complex traits:

- *Incomplete penetrance and phenocopy*. Some individuals who inherit a predisposing allele may not manifest the disease (incomplete penetrance), whereas others who inherit no predisposing allele may nonetheless get the disease as a result of environmental or random causes (phenocopy).
- *Genetic (or locus) heterogeneity*. Mutations in any one of several genes may result in identical phenotypes, such as when the genes are required for a common biochemical

pathway or cellular structure. For these reasons medical geneticists typically have no way of knowing whether two patients suffer from the same disease for different genetic reasons, at least until the genes are mapped. Genetic heterogeneity hampers genetic mapping, because a chromosomal region may cosegregate with a disease in some families but not in others. Genetic heterogeneity should be distinguished from allelic heterogeneity, in which there are multiple disease-causing mutations at a single gene. Allelic heterogeneity tends not to interfere with gene mapping.

- *Polygenic inheritance.* Some traits may require the simultaneous presence of mutations in multiple genes. Polygenic traits may be classified as discrete traits, measured by a specific outcome, or quantitative traits, measured by a continuous variable whose level may be set by the combined action of individual quantitative trait loci.

The methods available for genetic dissection of complex traits fall neatly into four categories: linkage analysis, allele-sharing methods, association studies in human populations, and genetic analysis of large crosses in model organisms such as the mouse and rat.

Linkage analysis involves proposing a model to explain the inheritance pattern of phenotypes and genotypes observed in a pedigree. It is the method of choice for Mendelian diseases, such as Huntington disease and cystic fibrosis, for which the diagnosis is less ambiguous and there is a near one-to-one correspondence between genotype and phenotype, however applications to complex traits can be more problematic.

Linkage analysis is the starting point for positional cloning. Families in which a disease phenotype segregates are analyzed using a group of DNA polymorphisms. The earliest and still most fully documented success in which linkage mapping alone led to the gene was cystic

fibrosis in 1989. At that time, the polymorphic markers were restriction fragment length polymorphisms; today hundreds of thousands of single nucleotide polymorphisms are available through the sequencing of the human genome.

The nature of linkage mapping limits resolution to the order of 1-10 cM. In favourable cases (such as cystic fibrosis), the pattern of crossovers in the region of the gene among the cohorts studied leaves only a few predicted genes, all within about 1 cM (~ 1 Mb), as likely candidates (Botstein and Risch, 2003).

Allele-Sharing methods involve studying affected relatives in a pedigree to see how often a particular copy of chromosomal region is shared identical-by-descent that is inherited from a common ancestor within the pedigree.

Association studies do not concern familial inheritance patterns at all, but they are case-control studies based on a comparison of unrelated affected and unaffected individuals from a population. Association studies are not well suited to whole-genome searches in large, mixed populations. Because linkage disequilibrium extends over very short distance in an old population, one would need tens of thousands of genetic markers to cover the genome.

There are some differences between linkage analysis studies and association studies. Associations studies test whether a disease and an allele show correlated occurrence in a population, whereas linkage studies test whether they show correlated transmission within a pedigree. Associations studies focus on population frequencies, whereas linkage studies focus on concordant inheritance.

Experimental crosses of mice and rat offer an ideal setting for genetic dissection of mammalian physiology. The power of experimental crosses is most dramatically seen in the ability to dissect quantitative traits into discrete genetic factors. Systematic quantitative trait locus (QTL) mapping has only recently become

possible with the construction of dense genetic linkage maps for mouse and rat.

In the past decade, quantitative trait locus (QTL) mapping has identified hundreds of chromosomal regions containing genes affecting asthma, atherosclerosis, diabetes, hypertension, obesity and other complex phenotypes. An important goal of QTL mapping is to identify the genes underlying these polygenic traits and to gain a better understanding of their physiology and biochemistry. Korstanje and Paigen in 2002 suggested that QTL mapping of complex traits is a promising technique and that the harvest of QTL genes is just beginning.

The analysis of QTL has also been made easier by recently developed statistical tools for detecting gene-gene interactions. Identifying QTL genes in animal models has more often relied on positional cloning followed by the use of transgenic or knockouts. QTL found in animal models may predict their location in the human genome.

The primary difficulty in identifying QTLs is that any single QTL explains only a small fraction of the phenotypic variation and thus the phenotype-genotype correlation is low. QTL analysis can be separated into two stages: (1) from phenotype and genotype to QTL, and (2) from QTL to gene. (Tab 2)

Stage 1 consists of correlating genetic markers (i.e. detectable DNA polymorphism, such as microsatellites and single nucleotide polymorphisms [SNPs]), with a phenotype in a segregating population. Therefore, stage 2 is required to assign the phenotype to a single gene (and not a locus) preferably by means of functional assay. (Darvasi and Pisanté-Shalom, 2002)

	From phenotype and genotype to QTL	From QTL to gene
Humans	Linkage analysis (affected sib pairs; extended families; discordant sib pairs; etc.)	Positional information
		Functional information
	Association analysis (case-control; TDT; etc.)	
Model organisms	Linkage and association analysis with experimental crosses (intercross; backcross; recombinant inbred strains; congenics; etc.)	Positional information
		Functional information Transgenics
	Linkage and association analysis with existent populations (half-sibs; full sibs; etc.)	Knockouts
		Reciprocal hemizygosity[a]

Table 2 Main stages and strategies in QTL mapping *From Darvasi and Pisanté-Shalom, 2002.*

1.2.1 Linkage disequilibrium (LD)

Mendelian diseases (e.g. cystic fibrosis) result from defects in one gene. However, it is suspected that most common diseases (e.g. hypertension) are 'polygenic', meaning that variation in the presence/absence of the disease is attributable, at least in part, to polymorphism of multiple interacting genes.

An allele of one of these 'susceptibility genes' may contribute to disease in certain genetic and environmental contexts, but not in others (Kirk *et al.*, 2002).

Polymorphism represents the transient state between the origin of genetic variation by mutation and the loss of variation by fixation of either the ancestral or derived state. Today, the expression 'single nucleotide polymorphism' (SNP) is often applied to variable sites for which the rarer base is present within the population at >1%

frequency, whereas germline polymorphism with a frequency <1% are typically referred to as 'mutations'

Traditional linkage analysis, based on allele sharing between relatives, identifies broad chromosomal regions that are likely to contain disease genes. However, the resolution of these methods is limited by the number of recombination events in typical pedigrees and is impractical for positional cloning efforts in complex disease. Fine mapping within the broad regions identified by allele-sharing methods is a major challenge. Gene mapping strategies based on linkage disequilibrium (LD) are expected to have much greater resolution, and should be able to capitalize on dense SNP maps as they become available. (Risch and Merikangas, 1996)

LD mapping can precisely locate genes of small effect and could be used to identify common disease genes in genomewide scans or to reduce the number of candidate genes in a region in which linkage has been established. (Lander 1996; Risch and Merikangas 1996; Collins et al. 1997; Lai et al. 1998). In the presence of common disease alleles, or when the frequency of rare alleles is increased through selection, the sample sizes required for LD studies are much smaller than those for equivalently powered linkage studies (Risch and Merikangas, 1996). Unlike traditional linkage studies, the power of LD mapping depends strongly on disease-allele frequencies and on the extent of disequilibrium between marker and disease alleles (Muller-Myhsok and Abel, 1997; Tu and Whittemore, 1999).

LD is the correlation of character states among polymorphic sites. (Lynch and Walsh, 1997 in Kirk *et al.*, 2002). The simplest explanation for LD is linkage coupled with insufficient passage of time to randomize the character states by meiotic recombination.

As ancestral haplotypes propagate through a population, their physical length is reduced by recombination events. Recombination events between markers separated by very short distances are very rare. Individuals inheriting a disease mutation from a common, but possibly distant, ancestor are expected to share a region of the

ancestral haplotype in which the mutation originated. Markers within this shared haplotype are non-randomly associated with the disease and each other, and are said to be in linkage disequilibrium.

Association studies using family based controls can distinguish between linkage disequilibrium and other sources of association, such as population admixture (Abecasis and Cookson, 2000).

Association studies using LD to identify susceptibility genes have achieved some recent success. Proof of principle experiments using a high-density SNP map in genomic regions already known to contain susceptibility genes for complex diseases have confirmed known genes or identified new ones (Rioux *et al.*,2001; Hewett *et al.*, 2002; Horikawa *et al.*, 2000 in Kirk *et al.*, 2002).

These successful association studies are 'candidate gene' approaches, where the location of the susceptibility gene is either already known or suspected. This experimental design is based on a two-step LD methodology (Roses, 2000 in Kirk *et al.*, 2002). In the first step, a low-density SNP map (i.e. 'picket fence') of the whole genome is used to identify a region(s) containing the potential susceptibility gene(s). Once the region is identified, a high-density SNP map is used to home in on the susceptibility gene within the region.

1.3 LUNG CANCER

Lung cancer is the leading cause of cancer death in the world (12.3% of all cancer), causing more than one million deaths each year. The incidence and mortality rates are highest in the United States and Europe and continue to increase in industrialized countries (Rivera, 2004). Its leading cause is tobacco smoke, a complex mixture containing many carcinogens that play a role in the genetic and epigenetic changes occurring in the respiratory epithelium during the multistage process of lung carcinogenesis. The

many lung cancer-specific carcinogens (including polycyclic aromatic hydrocarbons and nitrosamines) in the particular matter of tobacco smoke have to be metabolized before they are either secreted or can bind to DNA with the formation of adducts. DNA adducts may be repaired or lead to apoptosis. If they persist, miscoding mutations in key genes, such as p53 or RAS may cause genetic instability, leading to further mutational damage and eventually to cancer. (Fig. 2 in Minna *et al.*, 2002)

About 40% of all new lung cancers occur in former smokers, who remain at high risk of developing lung cancer for many years after they stop smoking.

A part from HIV/AIDS, tobacco is the only major cause of death that is increasing rapidly. Worldwide, smoking caused about three million deaths in 1990, out of a total of 30 million adult deaths from all causes, and it will cause about 10 million in 2030, out of a total of about 60 million (Peto *et al.*, 1999).

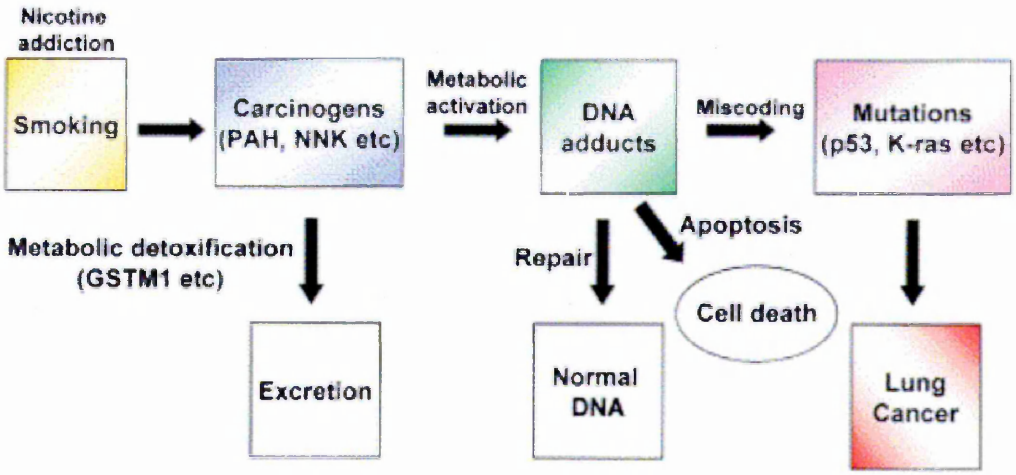

Figure 2 Nicotine addiction, tobacco smoke carcinogenesis, and the pathogenesis of lung carcinomas. *From Minna et al., 2002.*

Non-small cell lung cancer (NSCLC) accounts for 80 to 85% of all new cases of lung cancer. Less than 15% of patients with lung cancer survive their disease.

1.3.1 Histological Types

The term lung cancer is used to denote carcinomas arising from the epithelium of the respiratory tract.

Histopathogenesis

The histopathogenesis of lung cancer is incompletely understood and the true site of origin of lung cancer is unclear.

Two theories exist: (1) a pleuripotential cell (stem cell) theory, advocated by Auerbach, of one pleuriotential cell from which all cell types evolve, and (2) a small-cell theory, advocated by Yesner, of a type of small-cell neoplasm that undergoes transformation and evolves into the other cell types. (Fig. 3)

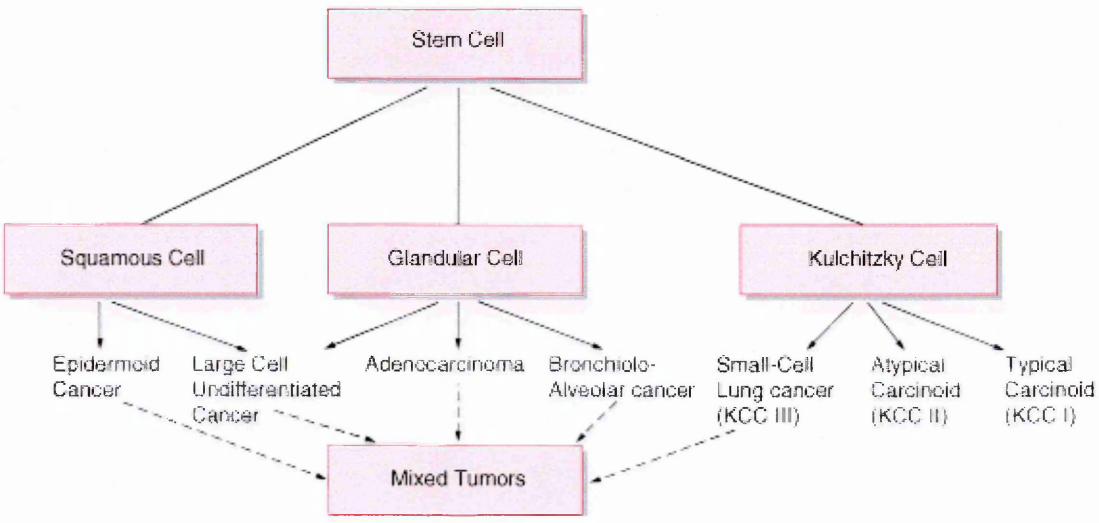

Figure 3 The stem cell theory of lung cancer pathogenesis as proposed by Auerbach.

The true site of origin of small-cell lung cancer is difficult to identify. It is also difficult to know the origin of adenocarcinomas or large-cell carcinomas because they may come from the bronchial epithelial surface or the underlying mucous glands. There is, however, increasing evidence that tumors arise from a common mucosal, pleuripotential stem cell. The heterogeneity found in various cell

types of lung cancer, particularly large-cell anaplastic carcinoma and various small-cell lung cancer tumors, may explain the variations in response to therapy and may be explained by the common stem cell theory.

Lung cancers arise after a series of molecular and morphological events. The molecular events commence in histologically normal epithelium and show a specific sequence (Hirsch et al. 2001 in Minna *et al.*, 2002). A microscopie based microdissection of epithelial tissue followed by allelotyping of smoking damage to lungs in lung cancer patients or current or former smokers without lung cancer revealed hundreds of thousands of lesions (~ 90,000 cells in size) containing clonal abnormalities of allele loss , occurring in both histologically normal as well as abnormal (e.g. hyperplasia, dysplasia or carcinoma in situ) respiratory epithelium (Park et al., 1999 in Minna *et al.*, 2002). (Fig. 4)

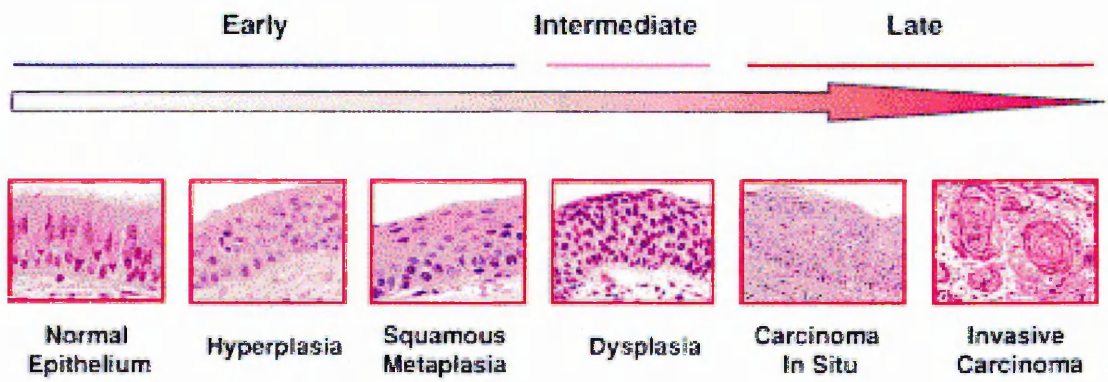

Figure 4 Progressive morphological and molecular changes during the multistage development of squamous cell lung carcinomas. *From Minna et al., 2002.*

In 1982, the World Health Organization classified bronchial carcinoma into four major histologic types (squamous cell carcinomas, adenocarcinoma, small cell carcinoma and large cell carcinoma), rare tumor types, and tumors consisting of more than

one histologic pattern. However, for clinical, therapeutic and biological reasons, lung cancers are classified into small cell (SCLC) and non-small cell (NSCLC) subdivisions.

Squamous cell carcinomas account for approximately one-third of all lung cancers. They usually originate within a central bronchus, and a central cavitation from necrosis is common. This kind of cancer generally grows slowly with little likelihood of distant spread.

Adenocarcinomas account for approximately 30% to 45% of lung cancers. These tumours originate within the periphery of the lung, in particular they arise from the progenitor cells of bronchioles (Clara cells) or alveoli (Type II pneumocytes) or from mucin producing cells, and tend to metastasize early to regional lymph nodes and to distant sites, particularly the brain. Adenocarcinoma is the most common lung cancer in the world, and its frequency is increasing rapidly especially in women, people who have never smoked, and in young people.
The term bronchioalveolar carcinoma is applied to those adenocarcinomas believed to arise from the terminal bronchioles and alveoli and represents approximately 2% to 4% of all lung cancers. Usually, the presence of these tumors is not correlated with smoking.

Large-cell carcinoma represents a less differentiated form of the other NSCLC (non-small cell lung cancer). This tumor accounts for approximately 9% of all lung cancers. The name derives from the large cells with large nuclei and prominent nucleoli that characterize the tumor. Large-cell carcinomas often have poor differentiation and no evidence of maturation. Giant- and spindle-cell carcinomas, that have been classified as a variant of large-cell carcinoma, carry poor prognoses and account for less than 1% of lung cancers.

Small-cell lung cancer (SCLC) accounts for 10% to 15% of all lung cancers. This kind of tumor has a very aggressive clinical course, with frequent widespread metastases. SCLC is not amenable to surgical treatment in its early stages. Small-cell lung cancers typically are situated in a peribronchial location, and they infiltrate the bronchial submucosa and peribronchial parenchyma tissues. SCLCs have neuroendocrine features and arise from cells programmed to differentiate along these lines.
Without treatment, small cell carcinoma of the lung has the most aggressive clinical course of any type of pulmonary tumor, with median survival from diagnosis of only 2 to 4 months.

Since the 1930s, cytologic examination of sputum has been used for the diagnosis of advanced and early lung cancer.
Squamous and SCLC arise mainly from the central airways, while adenocarcinomas (including bronchioloalveolar cancer) are peripherally located. Sputum samples contain exfoliated cells shed from the oropharynx and larger respiratory passages, so the cytologic analysis of sputum is helpful for the detection of central tumors arising from the larger exfoliated bronchi such as squamous cell and small-cell carcinomas. Cells from peripheral tumors arising from the smaller airways, such as adenocarcinomas, can be detected only occasionally in sputum sample (Gazdar and Minna, 1999). There is no squamous epithelium in the normal lung, and tumors arise from metaplastic changes resulting from smoking.

Genetic predisposition to developing lung cancer

Approximately 11% of tobacco smokers develop lung cancer. This suggests that some genetic factors may influence the risk for lung cancer among those who are exposed to carcinogens. (Fong *et al.*, 2005).
Tokuhata and Lilienfeld in 1963 and Ooi *et al.* in 1986 found increased risk for lung cancer in both smoking and non-smoking

relatives of lung cancer patients. It was found that smoking interacts with a family history of lung cancer. Familial aggregation studies of lung cancer provide indirect evidence for the role of genetic predisposition. More common genetic polymorphisms at genes, coding for enzymes involved with metabolism of exogenous compounds, including carcinogens, are more likely to have a large quantitative effect on the risk of lung cancer. The combination of two lung tumor associated polymorphisms results in an additive risk of lung cancer and this suggests genetic heterogeneity and polygenic inheritance of lung cancer risk.

Using candidate gene and genome wide approaches it has been hypothesised that lung cancers have accumulated 20 or more clonal genetic and epigenetic alterations as a multistep process. These genetic changes include activation of proto-oncogenes such as *myc*, *Kras*, *EGFR*, *Her-2/neu*, and *BCL-2* as well as the inactivation of TSGs (Tumor suppressor genes) like *p53*, *R b*, *FHIT*, and *p16*. Genetic alterations affecting TSGs usually involve two events: the loss of large chromosomal DNA regions of one parental Dna, and a smaller mutational event affecting the second allele. The early clonal genetic lesions that occur in smoking damaged preneoplastic bronchial epithelium are being identified, as are the molecular differences between small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLS), and between tumors with different clinical outcome.

1.4 MOUSE MODEL

Over the past century, the mouse has developed into the premier mammalian model system for genetic research. Scientists from a wide range of biomedical fields have gravitated to the mouse because of its close genetic and physiological similarities to humans,

as well as the ease with which its genome can be manipulated and analyzed.

Although yeasts, worms and flies are excellent models for studying the cell cycle and many developmental processes, mice are far better tools for probing the immune, endocrine, nervous, cardiovascular, skeletal and other complex physiological systems that mammals share. Like humans and many other mammals, mice naturally develop diseases that affect these systems, including cancer, atherosclerosis, hypertension, diabetes, osteoporosis and glaucoma. In addition, certain diseases that afflict humans but normally do not strike mice, such as cystic fibrosis and Alzheimer's, can be induced by manipulating the mouse genome and environment.

1.4.1 Mice for cancer studies

The inherited genetic determinants of cancer risk remain largely unknown and the identifications of cancer predisposing genes directly in humans is very difficult. Moreover, individual variation in the development of cancer in humans, even given identical exposures or identical germline changes in specific predisposing genes, offers the possibility of profound insight into host factors that modify the development of cancer. Determining those factors in humans is a daunting task (Klausner, 1999). The fact that mice can and do get cancer has fostered the important role of the mouse in cancer research.

These models are helping us to understand this disease as a complex genetic trait and thus to identify the multiple genetic variant alleles involved in pathways that affect individual cancer susceptibility. The mouse offers significant advantages for

identification of tumor susceptible genes and is useful for the study of complex influences of genetic background on tumor susceptibility.
A large portion of the mouse genome shows regions of homology and conserved syntenies with the human genome, and comparative genetic maps between mice and humans have been used to predict the location of human and murine disease genes on the basis of their mapping in other species. (Dragani *et al.*, 1995)
Compared to the total span of evolution, the lineage that gave rise to mouse and man diverged relatively recently. It was found that the essential functions of most genes, in general, and most, if not all, genes identified as important in human cancer, are likely to be very similar between mouse and man.
Furthermore, the use of mice as a model offers significant advantages:

- the size of the "families" for the analysis is restricted only by the dimensions of the animal house;
- the ability to control environmental exposure;
- the relatively low cost;
- the availability of dozens of inbred strains, outbred mice, recombinant inbreds, recombinant congenics, transgenic, knockouts, knockins, and even clones.

1.4.2 Mouse inbred strains

A useful mouse model for the study of cancer is mouse inbred strains. They are obtained by crossing brothers and sisters for at least 20 generations. They are homozygous at almost every genetic locus thus representing fixed, unique genotypes that can be repeatedly accessed as homogeneous populations, with defined allelic composition and predictable phenotypes.

Although some intermixing between strains has occurred, the inbred rodent strains differ genetically from each other and represent good animal models of human diseases. The distribution and frequency of different alleles, including those related to cancer predisposition, may be sufficiently represented in laboratory strains compared with the wild groups from which they were derived.

Many inbred strains of mice are available, they have a predominance of resistance or susceptibility alleles, and exhibit extreme sensitivity or resistance to tumor development in certain tissues.

Opportune crosses between susceptible and resistant inbred strains can give valuable information about the dominant or recessive nature of resistance, as well as providing indications about the number of the genes involved or their locations in the genome. (Balmain and Nagase, 1998)

1.4.3 Mice for the study of Lung cancer

Different mouse strains exhibit different propensities to cancers, whether they are spontaneous or induced.

Spontaneous lung cancer develops in approximately 3% of wild mice. The *Mus musculus* strains most sensitive to the spontaneous development of lung tumors include A/J and SWR; strains with intermediate sensitivity include Balb/c and O20; relatively resistant strains include CBA and C3H; and the most resistant strains include DBA and C57BL/6. (Lynch, 1926; Shimkin, 1955 in Tuveson and Jacks, 1999)

Some studies have revealed that tumor predisposition in different strains is controlled by multiple loci which control fundamental processes such as the tumor growth rate, ability to stimulate angiogenesis, or the risk of malignant progression (Hirst and Balmain, 2004). These different loci have been identified through

the evaluation of crosses between different inbred strains or species of mice (Tuveson and Jacks, 1999).

Mice susceptible to spontaneous lung tumors are also often sensitive to chemically-induced lung cancer. Induction of lung tumors with chemical carcinogens is very reproducible and almost invariably results in pulmonary adenomas and adenocarcinomas (Shimkin and Stoner 1975; Malkinson 1989). Furthermore, mice exposed to environmental carcinogens, develop tumours by a multistage process very similar to that seen in humans. Genetic alterations detected in mouse tumours involve genes (such as RAS, cyclin D1, RB1 (pRb), TP53 (p53) and CDKN2A (p16), that are the most commonly altered genes in human tumours.

Spontaneous lung tumors in mice are similar in morphology, histopathology, and molecular characteristics to human adenocarcinomas, but in contrast to human lung cancer, the majority of murine lung tumors are histologically early lesions such as hyperplasias and adenomas, which may originate from either Clara cells or alveolar type 2 cells that line alveoli (Thaete and Malkinson, 1991 in Tuveson and Jacks, 1999). These similarities underline the importance of the use of mice as a model for the study of human cancer.

In most cases murine lung tumors develop relative benign adenomas and rarely more aggressive adenocarcinomas and exhibit limited vascularisation and very few, if any, metastasize.

There is no relationship between the susceptibility of any given mouse strain to lung tumors and its susceptibility to tumors of other organs.

Mouse models for lung cancer can serve as a valuable tool not only for understanding basic lung tumor biology but also for the development and validation of new tumor intervention strategies as well as for identification of markers for early diagnosis.

Individual cancer risk is determined by combinations of resistance or susceptibility alleles inherited through the germline. The polygenic

nature of this inheritance pattern and the variable penetrance of the genes will, of course, mask any clear-cut familial clustering and complicate studies designed to find the loci involved by classical linkage analysis.

Genetic loci, which modify lung tumorigenesis, have been identified by crossing different inbred strain of known susceptibility and resistance to lung tumorigenesis.

The mouse inbred model is useful for statistical methods that take account of the fact that multiple genes make different quantitative contributions to the phenotype. Quantitative trait locus (QTL) mapping is an important method and during the last few years has accelerated the analysis of polygenic susceptibility to various diseases. Using QTL mapping, a large number of mouse tumor modifier loci have now been mapped that control the size, growth rate or number of lesions that develop, or can influence different stages of tumorigenesis (Balmain and Nagase, 1998).

Rare mutations or polymorphisms, which have major effects on tumour growth or survival, contribute to only a small fraction of tumours in the human population.

Highly penetrant mutations in genes such as *BRCA1/2* are responsible for a proportion of cancers that show familial aggregation, and mouse models of such genes have provided many fundamental and unexpected insights into tumorigenic processes. (Hirst and Balmain, 2004).

1.5 BACKGROUND AND AIM OF THE PROJECT

The present project aims to clone candidate genes for the Pulmonary adenoma susceptibility 1 (*Pas1*) locus, that represents the major quantitative trait locus (QTL) affecting inherited predisposition to lung tumorigenesis in mouse.

In our laboratory, by using (LD) analysis in 21 strains of known susceptibility to lung cancer and by assembling a YAC contig, we mapped the *Pas1* locus to a ~1.5-Mb region on distal mouse chromosome 6. Our results, based on haplotype and phenetic analysis, suggest that the $Pas1^s$ susceptibility allele is shared by several mouse-inbred strains of independent origin, which show either high or intermediate predisposition to lung tumorigenesis. We have previously mapped in the A/J mouse (a strain that is highly susceptible to lung tumorigenesis) the *Pas1* locus to the distal region of chromosome 6. (Gariboldi. et al. 1993). Independent studies with the A/J strain have confirmed the major role of *Pas1* in mouse lung tumorigenesis and have also supported the fact that the quantitative trait locus (QTL) peak for *Pas1* is localized around *Kras2*. (Devereux et al. 1994; Festing et al. 1994; Manenti et al. 1995) As in many QTL studies, the linked region is too wide (> 10 cM) to undertake positional cloning of the causative gene. We hypothesize that the $Pas1^s$ allele originated from a single founder and we designed a linkage disequilibrium (LD) study in mouse strains to test this hypothesis and to eventually narrow the candidate region for *Pas1.* Herein, it was reported that several markers located in the telomeric region of chromosome 6 exhibit significant LD with the genetic predisposition to lung cancer development. (Manenti *et al.*,1999).

According to the Ensemble database, five known genes map in this region: the branched-chain aminotransferase cytolic 1 (*Bcat1*), the lymphoid-restricted membrane protein (*Lrmp*), the unnamed 493046P12Rik transcript, the transforming protein p21 (*Kras2*) and

the unnamed 4933403M22Rik transcript. Protein sequence comparison by BLAST showed that the 4933403M22Rik gene encodes a protein containing an intermediate filament tail domain related to lamin A, therefore, it was provisionally named *Lmna-rs1* (lamin A-related sequence 1). The 493046P12Rik gene encodes a growth hormone-inducibile soluble protein with no known protein domains; the gene was provisionally named *Ghiso.*

Using a different program to identify coding regions in the genome (NIX analysis package), we identified and then cloned a new gene mapping between *Lrmp* and *Ghiso*. This gene resulted the mouse homologues to the human AK001783 transcript and corresponds to the recently identified *Casc1* gene (Zhang et al., 2003), except for an additional 70bp in the 5'-UTR. (Manenti *et al.*, 2004).

The present project aims to identify and characterize allelic variants of candidate mouse genes mapping in the restricted pulmonary adenoma susceptibility 1 (*Pas1*) locus region.

The milestones of this project were:

1. To narrow the *Pas1* containing region by: (a) discovering new simple sequences length polymorphism (SSLPs) and single nucleotide polymorphism (SNPs) in the region around *Kras2* gene; (b) LD analysis in 29 strains susceptible or resistant to lung tumorigenesis; (c) analysis of recombination of *Pas1* region in (CxB)RI mice.
2. To analyse genetic polymorphism of the candidate genes mapping in the minimal region in mouse inbred strains of known susceptibility (A/J) and resistance (C57BL/6J) to lung tumorigenesis.
3. To test functional activity of candidate genes in vitro, by subcloning them in to a mammalian expression vectors and transfecting them into human cancer cell lines in vivo for: (a) clonogenic assay experiment; (b) cell count experiment; (c) gene level expression.

Chapter 2 MATERIAL AND METHODS

2.1 ANIMAL AND TREATMENT

2.1.1 Mouse strains and DNAs

A panel of 29 genomic DNAs from mouse inbred strains, plus DNAs of 13 CxB RI lines were purchased from The Jackson Laboratory (Bar Harbor, ME, USA). C57BL/6J and A/J mice were also from The Jackson Laboratory.

2.1.2 Lung tumor induction

F1 hybrids (ABF1) of male C57BL/6J and female A/J crosses were treated intraperitoneally with urethane (BDH, Poole, UK) dissolved in water (1000 mg/kg body weight) at 4 weeks of age and observed until 50-60 weeks of age, when the mice were killed. Normal lungs from untreated A/J and C57Bl/6J male mice of about 6 months of age were used for RNA preparation. Lung tumors and normal tissue were excised, immersed in RNAlater (Qiagen, Valencia, CA, USA) and maintained at –20°C.

2.2 IDENTIFICATION AND ANALYSIS OF GENETIC VARIATIONS

In A/J and C57BL/6J mice, I carried out a nucleotide sequence analysis of genomic fragments for a total length of about 20000 bp in the region of mouse Chromosome 6 spanning *D6Int20* to *D6Int45*, corresponding to the region from 145.027489 to

146.035280 Mb according to Ensembl Mouse release 13.30.1 (http://www.ensembl.org/).

Genotype analyses were performed using the polymerase chain reaction (PCR) technique and two different kind of genetic markers were used:

SNP (Single Nucleotide Polymorphism)

SSLP (Single Sequence Length Polymorphism)

In all, 65 genetic markers were used in genotyping 29 inbred strains.

In the 29 strains, SNPs were analysed either by allele-specific oligonucleotide hybridization (ASO) or by direct sequencing, whereas SSLPs were analysed by polyacrilamide gel electrophoresis or by agarose gel electrophoresis.

2.2.1 SNP analysis

We identified SNPs in genes and in genomic DNA fragments.

All PCR primers used in this study were designed using the Primer 3 software (http://www.genome.wi.mit.edu) and purchased from MWG Biotech, Ltd (Ebersberg, Germany). To generate new SNPs, the nucleotide sequence of the chromosomal region was analysed by the NIX program to identify putative exons. PCR primers were designed in the genomic regions surrounding the putative exons and amplified in A/J and C57BL/6J parental strains.

SNP identification by sequence analysis

Sequencing was performing using an ABI PRISM automatic DNA sequencer (Applied Biosystem), and compared using the fragment Assembly System contained in the Genetics Computer Group (GCG)

software package (Wisconsin Package Version 10.2, Madison, WI, USA). When a polymorphism was found between the two sequences (SNP), two oligonucleotides were designed covering the region where the polymorphism was present, and these oligonucleotides were used to analysed the 29 DNAs by ASO hybridization.

ASO hybridization

Oligonucleotide hybridization has the potential to provide a method of detecting single-base changes within genomic DNA. (Corner *et al.*, 1982). Originally carried out on total genomic DNA, ASO is now used exclusively to establish the presence or absence of a specific mutation in PCR-amplified fragments of genomic DNA.

Probes are constructed from sequence spanning the variant nucleotide site. ASO probes are typically 15-20 nucleotides long and are normally employed under hybridization condition at which the DNA duplex between probe and target is stable only if there is a complementary perfect base between them. Typically, this involves designing the oligonucleotides so that the single nucleotide difference between alleles occurs in a central segment of the oligonucleotide sequence, thereby maximizing the thermodynamic instability of a mismatched duplex. For this work allele-specific 15-mer oligonucleotides empassing the SNP were 5'-end-labeled with [γ-^{32} P]dATP (3000 Ci/mM) (Amersham, Branchburg, NY) and T4 polynucleotide kinase (New England Biolabs, Beverly, MA).

The ASO procedure consists in spotting the denatured PCR-amplified fragment on to a nitrocellulose membrane and hybridised, in presence of tetramethylammonium chloride (TMAC), with the probes labelled with a radioactive isotope. After allowing sufficient time for probe-target heteroduplex formation, the probe solution is decanted, and the membrane is washed to remove excess probe that may have become non-specifically bound to the filter. It is then dried and exposed to an autoradiographic film.

2.2.2 SSLPs analysis

Simple sequence length polymorphisms consist of short tandem repeats whose units are generally 2-6 bp in length. They are known to distribute somewhat uniformly throughout the genome and to allow for a greater information content per locus than SNPs. (Dietrich *et al.*, 1992)

SSLPs are used to be the method of choice for genetic mapping in humans, mice and rats, due to their exceptionally high rate of polymorphism and their relative ease of use. In mouse and rat, the markers make it feasible to map the entire genome in any cross between laboratory strains. (Dietrich *et al.*, 1992)

In this work, SSLP markers were typed with PCR primers and 25 radioactive PCR cycles (55°C annealing temperature); results were scored on 6% denaturing polyacrilamide gels.

2.2.3 Fine-mapping recombination breakpoint in CxB RI line

(CxB) RI lines were derived from inbred Balb/c and C57BL/6J strains by intercrossing individuals from the F2 generation and then inbreeding for at least 20 generations of brother-sister mating. All mice of a particular RI line are genetically identical and homozygous at all loci, with approximately half the loci derived from C57BL/6J and half from Balb/c.

The genotypes of the distal region of Chromosome 6 of CXB RI strains (http://www.informatics.jax.org) were retyped for confirmation and compared with the lung tumor susceptibility data of the lines (Malkinson and Beer, 1984).

The CxB5 RI line was genotyped using 45 markers flanking the *Pas1* region, in order to precisely define the extension of the C57BL/6J-inherited genome around the *Kras2* gene.

2.3 *Pas1* CANDIDATE GENE IDENTIFICATION

The genomic region between markers *D6Int24* and *D6Int15* (*Pas1* core haplotype) was extensively analyzed to identify all putative coding regions based on both homology search and gene prediction program present in the NIX analysis package. The mRNA sequences of *Bcat1* (AK013888), *Lrmp* (NM_008511), *Ghiso* (AF412300), *Kras2* (NM_021284) and *Lmna-rs1* (AK016641) were retrieved from the Genebank. Computational sequence analysis techniques, BLAST and NIX, were used to assemble the *Casc1* gene (Acc. #AY485607, AY485608). The accuracy of this prediction was evaluated by PCR analysis of cDNA with primers located in the middle of putative exons. The 5′ end of the *Casc1* gene was determined by 5′ RACE using the Sure-RACE™ panels (OriGene Technologies, Rockville, MD, USA).

The sequences of the six genes were PCR-amplified from normal lung cDNA of A/J and C57BL/6J mice with the following primer sets: *Bcat1*, 5′-ATTCCAACGATGGAGAATGG-3′, 5′-TCTTCTGTGGCACCGTCAC-3′, *Lrmp*, 5′-CTGACACCCATCTCATGTGC-3′, 5′-GAAGGAAACGTTAAAATCCAAGG-3′, *Casc1*, 5′-CTCACCGGAAGACTACAGTGC-3′, 5′-GAACGCCTGGTGCTCTTAAA-3, *Ghiso*, 5′-TCTGAGCAAGCTGTCGTCTG-3′, 5′-TGGTGCAATGATTGGTCAG-3, *Kras2*, 5′-GGAGAGAGGCCTGCTGAAA-3′, 5′-AAGTGTGCCTTAAGAAAGAGTACAAA-3, *Lmna-rs1*, 5′-CTCTGCCAGCCATGATGAA-3′, 5′-TTTCATGGCTTTCCATTTTCTT-3. PCR products from the two mouse strains were sequenced and compared for polymorphism.

2.3.1 Allele-specific expression of candidate genes in mouse lung tumors

Normal and lung tumor samples were lysed in 700 µl of lysing solution using the MM 300 mixer mill (Retsch GmbH & Co., Haan, Germany). Total RNA was extracted according to the RNeasy Midi kit (Qiagen, Valencia, CA) protocol, including a DNase I treatment performed directly on the column (see below).

PolyA$^+$ mRNA was prepared by mixing total RNA with Oligotex suspension (Qiagen, Valencia, CA). Aliquots of polyA$^+$ mRNA were reverse-transcribed using oligo(dT) primers and Superscript™ for RT-PCR (Invitrogen, Carlsbad, CA). The first strand cDNAs were used for expression analysis of candidate genes by a semi-quantitative PCR, adjusting the number of PCR cycles to avoid saturation of PCR amplification. All reactions were performed in a final reaction volume of 12 µl containing 1X GeneAmp buffer, 1.5 mM $MgCl_2$, 5 pmoles of each primer, 200 µM of dNTPs and 0.6 U of AmpliTaq Gold (Applied Biosystems, Foster City, CA). Thermocycling was performed on the Gene Amp PCR System 9700 (Applied Biosystems). Primer sets were designed to obtain a PCR product 170-400 bp in size. The *Gapdh* gene was amplified with primers 5'-TGTTCCTACCCCCAATGTGT-3' and 5'-GTGGAAGAGTGGGAGTTGCT-3', producing a 179–bp fragment, and used to normalize the amount of cDNA in each sample. Candidate *Pas1* genes were RT-PCR-amplified with the following primers: *Bcat1*, 5'-ATTCCAACGATGGAGAATGG-3', 5'-TCTTCTGTGGCACCGTCAC-3, *Lrmp*, 5'-ACGCTGAGGACGAGAGAAGT-3', 5'-TGCTCTGTTGGCTCTTCTGA-3', *Casc1*, 5'-TGAAGCTGCTGAGTGAAGCA-3, 5'-ACTAGCACCCAGCCCTTCAC-3', *Ghiso*, 5'-TCATGTATAAAAGTGGATCAATTCC-3', 5'-TGGTGCAATGATTGGTCAG-3, *Kras2*, 5'-GGAGAGAGGCCTGCTGAAA-3', 5'- ACCATAGGCACATCTTCAG-3, *Lmna-rs1*, 5'-CCTATCCACTGGAAGCAAGC-3, 5'-TTTCATGGCTTTCCATTTTCTT-3'.

PCR products were loaded on 2.5% agarose gels, stained with ethidium bromide, and photographed with a CCD camera for image analysis and quantitation.

For allele-specific gene expression, RT-PCR products containing SNPs were sequenced with both the forward and reverse primers. The following primer sets were used: *Lrmp*, 5′-CATCGCTTCATCAGCACCTA-3′, 5′-CTCAGAAAGGCTGGGCTTC-3′, *Casc1*, 5′-GCAGGAGGAGGAGGAGAGAC-3, 5′-TGTTCAAAGGCCTGGTTCTT-3′, *Ghiso* 5′-TCATGTATAAAAGTGGATCAATTCC-3′, 5′-TGGTGCAATGATTGGTCAG-3, *Kras2* 5′-GGAGAGAGGCCTGCTGAAA-3′, 5′-ACCATAGGCACATCTTCAG-3. The same *Kras2* primers also allowed detection of *Kras2* mutations at codon 61 by nucleotide sequence analysis of the PCR product.

Sequence pherograms were analyzed and the heights of peaks at the SNP position were evaluated.

2.4 STATISTICAL ANALYSIS

According to their susceptibility to lung tumorigenesis, mouse strains were placed into two phenotype groups: the first group contained resistant strains and the second group contained susceptible and intermediate strains. Linkage disequilibrium between strain segregation of marker alleles and lung tumor susceptibility was evaluated by Fisher's exact test provided by the SAS software (SAS, 1998). *P* values were transformed to their negative logarithms, and a significant LD was considered if the $-\log P > 2$ ($P < 0.01$).

Statistical analysis of cell counts in our transfection, were performed

by univariate analysis of variance using SPSS 10.1 software (SPSS Inc., Chicago, IL).

2.5 CANDIDATE GENES ANALYSIS

2.5.1 Cell lines

Human non-small cell lung carcinoma-derived cell line A549 and NCI-H520 were purchased from American Type Culture Collection (Rockwill, MD). Cells were grown in RPMI-1640 medium (Cambrex) supplemented with 5% FCS, at 37°C in a 5% CO_2 humidified atmosphere.
Human embryonic kidney cells (HEK-293T) were maintained in Dulbecco's modified Eagle medium (Cambrex), supplemented with 10% FCS, at 37°C in a 5% CO_2 humidified atmosphere.

2.5.2 Cloning of allelic variants in a mammalian expression vector

Casc1, *Ghiso*, *Lrmp*, and *Lmna-rs1* full-length transcripts derived from normal lung tissue of A/J or C57BL/6 mice were PCR-amplified for subcloning in the eukaryotic pEF6/V5-His-TOPO cloning vector (Invitrogen). The following primers located 5′ and 3′ of the respective coding regions to include the natural ATG start codon and end just ahead of the TGA stop codon to produce a chimeric recombinant protein containing the V5His epitope at the C-terminus were used: *Casc1*, 5′-GTGAGCGTTCCTGCTCTGC-3′, 5′-GGAATAGCTGAGGACGCGGATA-3′; *Ghiso*, 5′-GTGAAAATGGCCAATTCGT-3′, 5′-GACTTTGGTATCTGAATAGTAACG-3′; *Lrmp*, 5′-CTCAGCCCATGCTCTGTGTA-3′, 5′-CACTGGCGGTGGTCCATCA-3′; *Lmna-rs1*, 5′-

CTCTGCCAGCCATGATGAA-3', 5'-TGACTTTGCTTTCTTTTTCTTGGTC-3'.

All reactions were performed in a final volume of 50 µl containing 1X GeneAmp buffer, 1.5 mM $MgCl_2$, 5 pmol of each primer, 100 µM of dNTPs and 1.2 U of AmpliTaq Gold (Applied Biosystems). Thermocycling was performed on the Gene Amp PCR System 9700 (Applied Biosystems).

Recombinant expression vectors were sequenced, and clones containing full-length inserts in the correct orientation and without PCR-induced mutations were used for cell transfections. DNAs of the selected recombinant plasmids were prepared and purified using the Plasmid Midi Kit (Qiagen).

2.5.3 RNA expression

A549 and NCI-H520 cells seeded in 60-mm dishes at $5x10^5$ cells/well were transiently transfected with 3 µg recombinant plasmid DNA and 6 µl SuperFect (Qiagen) according to the manufacturer's protocol.

Superfect Transfection Reagent is a specifically designed activated dendrimer possessing a defined spherical architecture, with branches radiating from a central core and terminated at charged amino groups. Superfect Reagent assembles DNA into compact structures, optimizing the entry of DNA into the cell. Superfect-DNA complexes possess a net positive charge, which allows them to bind to negatively charged receptors (e.g. sialylated glycoproteins) on the surface of eukaryotic cells. Once inside, Superfect Reagent buffers the lysosome after it has fused with the endosome, leading to pH inhibition of lysosomal nucleases. This ensure a stability of Superfect-DNA complexes and transport of intact DNA to the nucleus.

Cells transfected with the empty vector were used as control. The transfections were stopped after 24h, 48h and 72h, cells were pelleted and stored at –80°C. The pellets were used for RNA extraction by RNeasy Midi Kit (Qiagen, Valencia, CA). This technology combines the selective binding properties of a silica-gel-based membrane with the speed of spin technology. A specialized high-salt buffer system allows up to 1 mg of RNA longer than 200 bases to adsorb to the RNeasy silica-gel membrane. Samples were disrupted in buffer containing guanidine isothiocyanate and homogenized. Ethanol is added to the lysate, creating conditions that promote selective binding of RNA to the RNeasy silica-gel membrane. The sample is then applied to the RNeasy column. Residual DNA is removed using the RNase-Free DNase Set. Total RNA binds the membrane, contaminants are efficiently washed away, and high-quality RNA is eluted in RNase-free water. The RNA concentration and purity were determined by measuring the absorbance at 260 nm (A_{260}) in a spectrophotometer. A same quantity of RNA was used to isolate PolyA$^+$ mRNA by Oligotex mRNA Mini Kit (Qiagen). cDNA was synthesized from polyA$^+$ mRNA using SuperScript-III reverse transcriptase (Invitrogen, Carlsbad, CA).

First-strand cDNAs were used for expression analysis of candidate genes by semiquantitative PCR. All reactions were performed in a final reaction volume of 12 μl containing 1X GeneAmp buffer, 1.5 mM $MgCl_2$, 5 pmol of each primer, 100 μM of dNTPs, and 0.6 U of AmpliTaq Gold (Applied Biosystems, Foster City, CA). Thermocycling was performed using the Gene Amp PCR System 9700 (Applied Biosystems). Primer sets were designed to obtain PCR products of 140-400 bp in size: *Gapdh*: 5′-GAAGGTGAAGGTCGGAGTC-3′, 5′-GAAGATGGTGATGGGATTTC-3′; *LRMP*: 5′-GAGTGAACACCGTCCCTCAT-3′, 5′-AAAGACTTGGCTTCCTTGTCC-3′; *Casc1*: 5′-TATTTACTGAGATTCAAATACAAATTAAGG-3′, 5′-AGGGCCATCTGTCGATAAGC-3′; *GHISO*: 5′-GCGTGCCGGAAAGTATGTTA-3′, 5′-AGTAGGGCAGCTCAGGGATT-3′;

and *LMNA-RS1*: 5′-AGAAGGAAATCCCACCAACC-3′, 5′-TTGAGTTGACTGCTTATTGCTTTT-3′. PCR products were loaded on 2.5% agarose gels, stained with ethidium bromide, and photographed with a CCD camera for image analysis and quantitation.

2.5.4 Colony formation assays

We performed a kill curve to determine the appropriate concentration of Blasticidin to use for the selection of resistant cells. A549 and NCI-H520 cells were seeded in 12-multi-well plates at $1.2x10^4$ cells/well and different concentrations of antibiotic were added to the medium. The kill curve revealed that the concentration of Blasticidin is different for the two cell lines in particular it is 5μg/ml for A549 cells and 2 μg/ml for NCI-H520 cells.
A549 and NCI-H520 cells seeded in 6-multi-well plates at $3x10^5$ cells/well were transfected with 2μg recombinant plasmid DNA and 4 μl SuperFect (Qiagen) according to the manufacturer's protocol.
Cells transfected with the empty vector were used as a control. Two days after transfection, the appropriate concentration of antibody was added and transfectant clones were selected after 17 days for A549 and 26 days for NCI-H520 cells. Blasticidin-resistant clones were either fixed with methanol (100%) and stained with 10% Giemsa, or treated with trypsin, pelleted by centrifugation, resuspended in RPMI-1640 medium and counted using a Burker chamber. We repeated at least three independent experiments.

2.5.5 Protein and apoptosis analysis

HEK-293T cells in 100-mm dishes were transiently transfected with 5μg recombinant plasmid DNA using the calcium phosphate

precipitation method. After 24, 48 and 72 h, cells were lysed in RIPA buffer supplemented with 1/100 protease inhibitor cocktail (Sigma) for 20 min at 4°C. Lysates were centrifuged at 13000 rpm for 10 min at 4°C and protein content in the supernatant was measured using the Bradford method (BioRad). Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes (Hybond-C Super; Amersham Pharmacia, UK). Membranes were blocked in Tris-buffered saline containing 1% Tween-20 (TBS-T)/2% BSA for 1 h, washed 3 times for 10 min each with TBS-T at room temperature and incubated overnight at room temperature with anti-V5 mouse monoclonal antibody (Invitrogen) diluted 1:2500 in TBS-T/2% BSA. Apoptosis was analyzed in HEK-293T, A549, and NCI-H520 cells transiently transfected as described above, hybridized in Western blots overnight at room temperature with anti-human poly(ADP-ribose) polymerase (PARP) purified mouse monoclonal antibody (BD Biosciences) diluted 0.5 mg/ml in TBS-T/2% BSA. Blots were washed with TBS-T and incubated for 1 h with secondary antibody (sheep ECL anti-mouse IgG, peroxidase-linked species-specific whole antibody; Amersham Biosciences) diluted 1:4000 in TBS-T/2% BSA. After washing, blots were developed using ECL Plus (Amersham Biosciences).

2.5.6 Immunofluorescence microscopy

HEK-293T cells grown on 13-mm coverslips were transiently transfected with calcium phosphate, washed in PBS, and fixed in 4% paraformaldehyde for 5-10 min. After several washes in PBS, cells were permeabilized by incubation in 1% BSA, 0.2% Triton X-100 in PBS for 30 min, and incubated in 10% anti-goat serum for 30 min followed by anti-V5 mouse monoclonal antibody (Invitrogen) diluted 1:150. Cells were washed several times with PBS-0.2% Triton X-100, incubated with secondary antibody (Alexa Fluor 488 conjugate;

Cambrex) diluted 1:250, washed again in PBS-0.2% Triton X-100, and coverlips were mounted with Vectashield mounting medium including DAPI (DBA Italia). Fluorescent cells were detected using a Nikon Eclipse E1000 fluorescence microscope and photographed with a DXm1200F digital camera.

Chapter 3 RESULTS

The first aim of this project was to discover and to test new polymorphisms (both SSLPs and SNPs) in the *Pas1* locus region, around *Kras2* gene, in a population of 29 laboratory mouse inbred strains with known susceptibility to lung carcinogenesis. We detected a shared A/J- and C57BL/6J-type haplotype containing genomic as well as coding polymorphisms in six genes, supporting the hypothesis that *Pas1* is a multigene locus.

3.1 *Pas1* CONTAINING REGION

3.1.1 Physical delimitation of the *Pas1* locus

Identification of *Pas1* candidate genes by positional cloning requires the reduction of the mapping region to a manageable size. After the first mapping studies that placed the *Pas1* locus in a 10-20 cM interval (Gariboldi *et al.*, 1993), linkage disequilibrium (LD) analysis in mouse inbred strains of known susceptibility/resistance to urethane-induced lung tumorigenesis placed the *Pas1* locus in a ~1.5-Mb region (Manenti *et al.*, 1999).

In order to further shorten the region containing the *Pas1* locus we analyzed the (CxB)RI 5 mouse line. We decided to include this line for our analysis because, although this line has inherited a large segment of the telomeric portion of chromosome 6 from the susceptible BALB/c parental strain, it is however highly resistant to lung tumorigenesis developing 0.07 tumors per mouse after urethane treatment (Malkinson and Beer, 1984). This discrepancy was explained by the finding that the (CxB)RI 5 mouse carries the *Kras2* allele from the resistant C57BL/6J parent line indicating the presence of small recombination that replaced the *Pas1* susceptible

allele (from BALB/c) with the *Pas1* resistant allele (from C57BL/6J). Therefore, in this mouse line the extension of C57BL/6J inherited region around the *Kras2* gene would be compatible with the physical map of *Pas1*.

We carried out an extensive genotype analysis of this RI line genotyping 45 markers on chromosome 6 (Tab.3) and we localized the centromeric point of recombination between markers *D6Mit198* and *D6Int152* (a SSLP mapping at 142.931580 Mb). The telomeric end of recombination was localized in a 9.9-kb fragment between markers *D6Int47* (145.814802 Mb) and D6Int56 (145.824723). *D6Int47* is located in the middle of exon 4 of the *Lmna-rs1* gene, whereas *D6Int56* is located in the intronic region between exons 2 and 3 of the same gene.

This finding enabled us to further narrow the *Pas1* region to 468 kb. (Fig. 5)

Five known genes map in this 468-kb region: the branched-chain aminotransferase cytosolic 1 (*Bcat1*), the lymphoid-restricted membrane protein (*Lrmp*), the unnamed 4930469P12Rik gene that encodes a growth hormone-inducible soluble protein (AF412299) with no known protein domains, (the gene was provisionally named *Ghiso*: growth hormone-inducible soluble protein), the transforming protein p21 (*Kras2*) and the unnamed 4933403M22Rik gene that encodes a protein containing an intermediate filament tail domain related to lamin A. Therefore, the 4933403M22Rik gene was provisionally named *Lmna-rs1* (lamin A-related sequence 1). (Fig. 5).

	C x B line												
Marker	1	2	3	4	5	6	7	8	9	10	11	12	13
D6Mit86	C	B	C	*	C	B	B	C	C	C	B	*	B
D6Mit138	C	B	C	B	C	B	B	C	C	C	B	B	B
D6Mit83	C	C	C	B	C	B	B	B	C	C	B	B	B
D6Mit1	C	C	C	B	C	B	B	B	C	C	B	B	B
D6Mit50	C	C	C	B	C	B	B	B	C	C	B	B	B
D6Mit223	B	B	C	B	C	C	C	B	C	B	B	B	B
D6Mit166	C	B	C	B	C	B	B	C	C	C	B	C	B
D6Mit159	C	B	C	B	C	B	B	B	C	C	B	B	B
D6Mit268	B	B	C	C	C	C	C	C	C	C	B	B	B
D6Mit74	B	B	C	B	C	C	C	B	C	B	B	B	B
D6Mit274	B	B	C	B	C	C	C	B	C	B	B	B	B
D6Mit184	B	B	C	B	C	C	C	C	B	B	B	B	B
D6Mit316	B	B	C	C	C	C	C	B	B	B	B	B	B
D6Mit102	C	B	C	C	C	B	C	C	C	B	C	B	B
D6Mit54	C	B	C	C	C	B	C	C	C	C	C	B	B
D6Mit25	B	B	C	C	C	B	C	C	C	.	C	C	C
D6Mit198	B	B	C	C	C	B	C	B	C	C	C	B	C
D6Int24		C	C	C	B	C	B	C	C	C	C	B	C
D6Int28				C	B	C	B						
D6Int26					B								
D6Int27					B								
D6Int18					B								
D6Int28					B								
D6Int29	.	C	C	C	B	C	B	C	C	C	C	B	C
D6Int3	.	C	C	C	B	C	B	C	C	C	C	B	C
D6Int31	.	C	C	C	B	C	B	C	C	C	C	B	C
D6Int6	.	C	C	C	B	C	B	C	C	C	C	B	C
D6Int36	.	C	C	C	B	C	B	.	C	C	C	B	C
D6Int9	.	C	C	C	B	C	B	C	C	C	C	B	C
D6Mit57	B	C	C	C	B	C	B	C	C	C	.	B	C
Kras2_MALKINSON	B	C	C	C	B	C	B	.	.	.	.	.	.
Kras2_37bp	B	C	C	C	B	C	B	C	C	C	C	B	C
D6Mit14	B	C	C	C	B	C	B	C	C	C	C	B	C
D6Int14	.	C	C	C	B	C	B	C	C	C	C	B	C
D6Int46	.	C	C	C	B	C	B	C	C	C	C	B	C
D6Int47	.	C	C	C	B	C	B	C	C	C	C	B	C
D6Int56	.	.	.	C	C	.	B	.	.	.	.	.	.
D6Int153	.	.	.	C	C	.	B	.	.	.	.	.	.
D6Int48	.	C	C	C	C	C	B	C	C	C	C	B	C
D6Int15	.	C	C	C	C	C	B	C	C	C	C	B	C
D6Mit294	B	C	C	.	C	C	B	C	C	C	C	B	C
D6Mit201	B	C	C	C	C	C	B	C	C	C	C	B	C
D6Mit15	B	C	C	C	C	C	B	C	C	C	C	B	C
D6Mit26	B	C	C	C	C	C	B	C	C	C	C	B	C
D6Mit372	B	C	B	C	C	C	B	C	C	C	C	B	C

Table 3 Single nucleotide polymorphism used for the analysis of (CxB)RI line (the lines are indicated on the top of the figure). On the left are listed the markers used in this study. The centromeric point of recombination is located between markers D6Mit198 (in green) and D6Int152 (an SSLP not shown), while the telomeric end of recombination is located in a 9.9kb fragnment between markers D6Int147 and D6Int56 (in yellow). B is for Balb/c, C is for C57/BL6J

Using different programs to identify coding regions in the genome (NIX analysis package), we identified and then cloned a previously

uncharacterized gene mapping between *Lrmp* and *Ghiso*. This gene is the mouse homolog of the human AK001783 transcript and corresponds to the recently identified *Las1* gene (Zhang *et al.*, 2003) except for an additional 70 bp in the 5'-UTR. This new gene has 16 exons, spans a 36-kb region, and is flanked at its 3'-end by the 3'-end of *Lrmp* and at its 5'-end by the 5'-end of *Ghiso.*

Subsequent LD analysis of outbred mouse lines susceptible and resistant to both inflammatory response and lung tumorigenesis delineated a 0.45-Mb region around the *Kras2* gene, included in the previous 1.5-Mb region (Maria *et al.*, 2003).

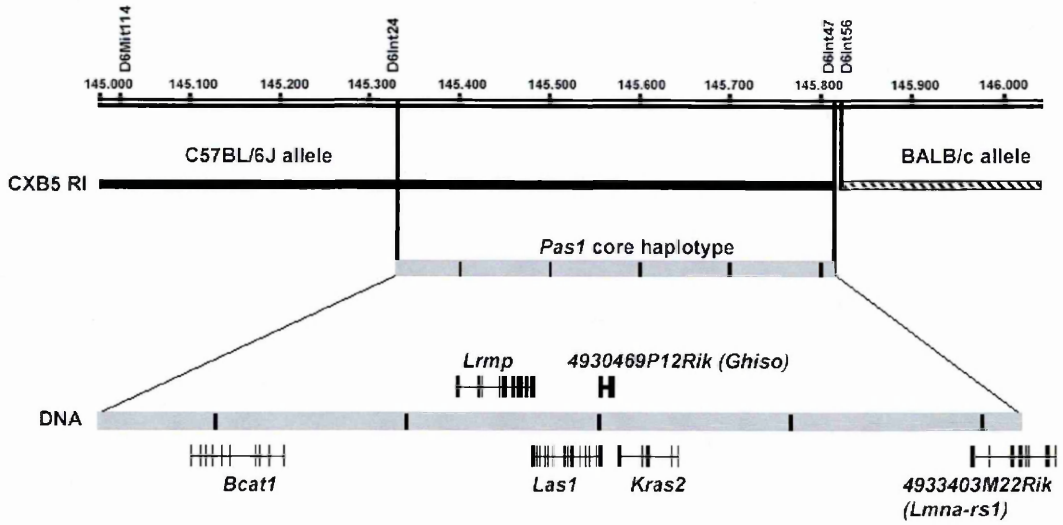

Figure 5 Physical map of distal Chromosome 6 and delimitation of the *Pas1* core haplotype. The upper line shows the genomic region from 145.027489 to 146.035280 Mb, and the point of genetic recombination in the mouse CXB5 RI line (black and hatched boxes indicate the C57BL/6J- and BALB/c- derived chromosomes, respectively). The genotyping of 65 genetic markers in this 1 Mb region identified a core *Pas1* haplotype, between D6Int24 and D6Int47, which contained 54 of these markers and the six genes shown. The approximate locations of the genes are depicted above (sense) and below (reverse) the segment representing the DNA sequence of that region. Introns are depicted by lines, the exons are depicted by solid vertical bars. Positions of the alignments are based on the Ensembl database. (*Las1* is for *Casc1*)

We performed a nucleotide sequence analysis of genomic fragments for a total of 20000 bp in the *Pas1* region and this allowed us to

identify 159 SNPs/SSLPs (Tab.4). Among these markers, we analyzed 27 SSLP and 37 SNP (see Tab.6 and Tab.7 for more details) between A/J and C57BL/6J mice selected according to their position and the distance between each other. We started our analysis genotyping these selected markers in 24 mouse inbred strains with known susceptibility to lung carcinogenesis (Tab.5). Some SNP markers were analyzed in 89 inbred strains with known susceptibility to lung carcinogenesis for a more complete analysis (see Fig.6 for ASO hybridization example).

Marker	Gene	Type	PCR Primers Name	Detection method	SNP Polymorphism	Amino-acid change
D6Int152		SSLP	CCTCTAGCCTCCTTGGGAAC TTCTGTCCTTTGGGGAAGG			
D6Mit114		SSLP	TGCAGAGACTCACAATTCACG AGGCCCTGTGCTGGCTAAG	Agar	145022	
D6Int20		SSLP	TGCAGAGACTCACAATTCACG AGGCCCTGTGCTGGCTAAG	32P		
D6Int16		SSLP	TGGGCCATGCTACAAGTTCT CCATGTTTGCTCTTGTGCTG	agar		
D6Int21		SSLP	TGGAGTGCTCTGAGCTGTTG CCTATCTGTCAGTTTTATCCCTTGT	agar		
D6Int22		SSLP	TCTGTACCTAGGGGCTGTGAA CACACCTGCATACAAGCACA	agar		
D6Int23		SSLP	GGAAGTAACTTGTGCCAGGATT TGCCCTAGGGAGTGTCTTATG	32P		
D6Int17		SSLP	CAATGCGGGTTTATCTGTCA TACATGCACGACCCTCACAT	agar		
D6Int24		SSLP	GGGAAATTGGTGGGAAGAAG CCTTGACTTTTGTTTCCTACCC	agar		
D6Int25		SSLP	TGTCTAACTTCCCAGACCAACA CAAAGTGGCAGTGTCCTCAA	agar		
D6Int26		SSLP	GGCATCTCATTCCTTCTTTCA CTGCTTGAATGCTTGGGTTT	agar		
D6Int27		SSLP	TGCATTCGTGTGCACATTC GTGCCCTTACCCACTGTCAT	agar		
D6Int18		SSLP	ATCCCACAACCTCATTCCAA CCTCCATTCTGTATGGCAAGA	32P		
D6Int28		SSLP	TGCTCACTGGCAACATGAAC GAGGCTCTCTCATGGCCTTT	agar		
D6Int19		SSLP	GGCCCATCCAGTCTTAGGTT CAGGTTGGCTTGGCTTATCT	32P		
D6Int29		SSLP	CCCTGATACCACACATGCAC CTAGCTGGGCCAAGAAAGC	32P		
D6Int30		SNP	GGGAGACAAACACACCTGCT GCAAGTGGATGCTAGGAACAG		TTTTACA[C/T]TGAAACC	Genomic
D6Int3		SSLP	GCCATCTCAGTGGAGGCTTT CATGTCAGAGATGAGCTGTGC	33P		
D6Int31		SSLP	AAACACGATATAGGCCCATTTAATA GTTGCTTTTCTTGGGCTGGT	33P		
D6Int53	*Lrmp_Ex1*	SNP	TGGCACCCTCTTGTCACGT CCCATGCTCTGTGTAAA		GTAAAAG[G/A]TCCCCCA	G -> N
D6Int4	*Lrmp_Ex2*	SNP	TTTTGCAGAAAGCCACTCTTT TCCTAGGACGCCCCATACTA		ACGGAAG[G/A]CTCCATC	G -> D
D6Int5		SSLP	TGAGCAGTCTCTCCAACCTCA TCTGCCAGTACCCACGTACA	33P		
D6Int32a	*Lrmp_Ex3*	SNP	TACGTGCCTAAAGCGACCTT TTTGTTCCCACTGGCTTTTC		GGTGCTG[A/G]CCACTT[G/C]TATTC[C/T]TGTACCA	D -> G

D6Int32b	*Lrmp*_Ex3	SNP	TACGTGCCTAAAGCGACCTT TTTGTTCCCACTGGCTTTTC		GGTGCTG[A/G]CCACTT[G/C]TATTC[C/T]TGTACCA	L -> F
D6Int32c	*Lrmp*_Ex3	SNP	TACGTGCCTAAAGCGACCTT TTTGTTCCCACTGGCTTTTC		GGTGCTG[A/G]CCACTT[G/C]TATTC[C/T]TGTACCA	S -> S
D6Int33		SNP	TACGTGCCTAAAGCGACCTT TTTGTTCCCACTGGCTTTTC		CGAGAGG[T/C]GGCGCCC	Intronic
D6Int34		SSLP 20bp_ins	GGTGGAATGCCTTGATCATT AGGAGCTGAGCTGCAGAAAG			
D6Int58	*Lrmp*_Ex6	SNP	AAATTCAGAGGGACAGATCCTACA GCTGCTTCTTCCTCTTTTGATTG	agar	T(A,B6/C(CBA)	CODING
D6Int35		SSLP	TCTTCCTGGCTCTGATGACC GGGCAATATTGGCTAAATATAAATCA	33P		
D6Int59	*Lrmp*_Ex11	SNP			TC[G/A]	S -> S
D6Int60	*Lrmp*_Ex11	SNP			AT[C(a=b6/A(cba)]	I -> I
D6Int6		SSLP	TCCAAAATGGCCCTGCTTAT GTTGCCCTCTGACTTCCACA	33P		
D6Int61	*Lrmp*_Ex13	SNP			AA[G/A]	K -> K
D6Int36		SSLP	GGGTGCTGTTGCTTAGTGGT ACACGTTTGTGGTGGTGTGA	33P		
D6Int54	*Lrmp*_Ex17	SNP	GTGGGTCTCACCTGACCATT CCTGCCCTCAGTAAGCACA		AGTGGAG[G/A]GGACCAG	G 438 R
D6Int7	*Lrmp*_Ex17	INS SSLP	GTGGGTCTCACCTGACCATT CCTGCCCTCAGTAAGCACA	agar		STL ->STLSTL
D6Int62	*Lrmp*_Ex18	SNP			GC[C(A,B6)/G(CBA)]	A -> A
D6Int8	*Lrmp*_Ex18	SNP	GAGCTTCCTCACAGGTCAGC TGCTGAGGACCCTAGCTTTG		GGACCAC[T/C]GCCAGTG	L 537 P
D6Int141	*Casc1*_Ex16	SNP	GGACCCCACCTAGGGTATTT GCAATCCCAGAATTCACGTT		CCTCCCC[C/G]GTGGCAA	P 700 P
D6Int154	*Casc1*_Ex16	SNP	GGACCCCACCTAGGGTATTT GCAATCCCAGAATTCACGTT		AGGGCAC[C/G]GAGTTCC	T 684 T
D6Int63						
D6Int64	*Casc1*_Ex16	SNP	GGACCCCACCTAGGGTATTT GCAATCCCAGAATTCACGTT		GAAGAAA[C/T]AGAGCAC	T 649 I
D6Int65	*Casc1*_Ex15g	SNP	CCACACAGACACTCCCACAC CCTTGAAACGAACCCTTCTG		CCTTCTC[G/C]GGTGTTT	Intronic
D6Int37a	*Casc1*_Ex14g	SNP	GACAATGCACAGCAGTGACG CCAGAGGGAGCTTGGAAAGT		AAGAAGA[G/T]AAAGGAA	Intronic
D6Int37b	*Casc1*_Ex14g	SNP	GACAATGCACAGCAGTGACG CCAGAGGGAGCTTGGAAAGT		TCTTGGG[A/C]CTCTTCT	Intronic
D6Int66	*Casc1*_Ex13g	SNP	CTGGGGACACACTGCCTAAC GGCTAATAGGCGCCTGAAAT		CCTAACT[A/G]TGGGCGG	Intronic
D6Int67	*Casc1*_Ex13	SNP	CTGGGGACACACTGCCTAAC GGCTAATAGGCGCCTGAAAT		ACCTTCA[G/A]TATGGAT	S 498 N
D6Int68	*Casc1*_Ex12g	SNP	TGCAGTTGTTTGCACTACCC CGCAGGAAGGTCTGAGGTTA		C/T	Intronic
D6Int69	*Casc1*_Ex12g	SNP	TGCAGTTGTTTGCACTACCC CGCAGGAAGGTCTGAGGTTA		C/G	Intronic
D6Int70	*Casc1*_Ex11g	SNP	GGCTGTATCAAGGGCTGACT TATCCTTCTTGACCGCAACA		C/T	Intronic
D6Int71	*Casc1*_Ex11g	SNP	GGCTGTATCAAGGGCTGACT TATCCTTCTTGACCGCAACA		A/G	Intronic
D6Int72	*Casc1*_Ex11g	SNP	GGCTGTATCAAGGGCTGACT TATCCTTCTTGACCGCAACA		A/G	Intronic
D6Int73	*Casc1*_Ex11g	SNP	GGCTGTATCAAGGGCTGACT TATCCTTCTTGACCGCAACA		G/A	Intronic
D6Int74	*Casc1*_Ex11g	SNP	GGCTGTATCAAGGGCTGACT TATCCTTCTTGACCGCAACA		G/A	Intronic
D6Int75	*Casc1*_Ex11	SNP	GGCTGTATCAAGGGCTGACT TATCCTTCTTGACCGCAACA		CACAGC[T/A]G[T/C]GCCTGG	C -> R
D6Int76	*Casc1*_Ex11	SNP	GGCTGTATCAAGGGCTGACT TATCCTTCTTGACCGCAACA		CACAGC[T/A]G[T/C]GCCTGG	L -> Q
D6Int77	*Casc1*_Ex12	SNP	GGCTGTATCAAGGGCTGACT TATCCTTCTTGACCGCAACA		C/T	Intronic
D6Int78	*Casc1*_Ex10g	SNP	ATGAGGCAGGTTACCAATGC TACGGACCAATACCCTCCTG		C/T	Intronic
D6Int79	*Casc1*_Ex9g	SNP	GCCAGCTTGGGCTACCTATC GGGGCATTAAAAAGGACACC		G/A	Intronic
D6Int80	*Casc1*_Ex9	SNP	GCCAGCTTGGGCTACCTATC GGGGCATTAAAAAGGACACC		AAAAG[C/T]TCTTTACT	L -> F
D6Int81	*Casc1*_Ex9	SNP	GCCAGCTTGGGCTACCTATC GGGGCATTAAAAAGGACACC		AAGGCC[A/G]TGACTGAA	M -> V

D6Int38	*Cascl*_Ex9	SNP	GCCAGCTTGGGCTACCTATC GGGGCATTAAAAAGGACACC		GAGCAAAA[T/A]CA[G/A]AACCCC	Q 278 Q
D6Int82	*Cascl*_Ex9	SNP	GCCAGCTTGGGCTACCTATC GGGGCATTAAAAAGGACACC		GAGCAAAA[T/A]CA[G/A]AACCCC	N -> K
D6Int83	*Cascl*_Ex9	SNP	GCCAGCTTGGGCTACCTATC GGGGCATTAAAAAGGACACC		ATTGCT[C/T]GTGGCCAG	T -> T
D6Int84	*Cascl*_Ex8g	SNP	GCCAGCTTGGGCTACCTATC GGGGCATTAAAAAGGACACC		C/T	Intronic
D6Int85	*Cascl*_Ex8g	SNP	GCACGCCTGCTCTAGACACT CCAGCTTCCTCCAGGAGTTC		T/G	Intronic
D6Int39	*Cascl*_Ex6g	SNP	CACTCACCAGACAGCAGCAC TGGTGGAAGGAGTGAAATGG		A/T	Intronic
D6Int86	*Cascl*_Ex6g	SNP	CACTCACCAGACAGCAGCAC TGGTGGAAGGAGTGAAATGG		G/A	Intronic
D6Int87	*Cascl*_Ex6g	SNP	CACTCACCAGACAGCAGCAC TGGTGGAAGGAGTGAAATGG		G/A	Intronic
D6Int88	*Cascl*_Ex6g	SSLP	CACTCACCAGACAGCAGCAC TGGTGGAAGGAGTGAAATGG	agar	5bp Del	Intronic
D6Int89	*Cascl*_Ex6g	SNP	CACTCACCAGACAGCAGCAC TGGTGGAAGGAGTGAAATGG		G/A	Intronic
D6Int90	*Cascl*_Ex6g	SNP	CACTCACCAGACAGCAGCAC TGGTGGAAGGAGTGAAATGG		A/G	Intronic
D6Int91	*Cascl*_Ex6g	SNP	CACTCACCAGACAGCAGCAC TGGTGGAAGGAGTGAAATGG		T/G	Intronic
D6Int92	*Cascl*_Ex6g	SNP	CACTCACCAGACAGCAGCAC TGGTGGAAGGAGTGAAATGG		G/T	Intronic
D6Int93	*Cascl*_Ex6g	SNP	CACTCACCAGACAGCAGCAC TGGTGGAAGGAGTGAAATGG		G/A	Intronic
D6Int94	*Cascl*_Ex6g	SNP	CACTCACCAGACAGCAGCAC TGGTGGAAGGAGTGAAATGG		T/G	Intronic
D6Int40	*Cascl*_Ex6g	SNP	ATCTCACCAGCCTGGAGCTT TTAATTCCTGCAGCCACCAC		C/T	Intronic
D6Int95	*Cascl*_Ex5g	SNP	ATCTCACCAGCCTGGAGCTT TTAATTCCTGCAGCCACCAC		C/A	Intronic
D6Int96	*Cascl*_Ex5g	SNP	ATCTCACCAGCCTGGAGCTT TTAATTCCTGCAGCCACCAC		A/C	Intronic
D6Int97	*Cascl*_Ex4g	SNP	AAACCCAAGCAAACAGACCA GCAAAAGGATGCCTTCGACAA		T/C	Intronic
D6Int98	*Cascl*_Ex4g	SNP	AAACCCAAGCAAACAGACCA GCAAAAGGATGCCTTCGACAA		T/C	Intronic
D6Int99	*Cascl*_Ex4g	SNP	AAACCCAAGCAAACAGACCA GCAAAAGGATGCCTTCGACAA		A/T	Intronic
D6Int100	*Cascl*_Ex4g	SNP	AAACCCAAGCAAACAGACCA GCAAAAGGATGCCTTCGACAA		C/T	Intronic
D6Int41	*Cascl*_Ex4	SNP	AAACCCAAGCAAACAGACCA GCAAAAGGATGCCTTCGACAA		AAGTGG[A/G]TCTGCTG	S -> N
D6Int101	*Cascl*_Ex4	SNP	AAACCCAAGCAAACAGACCA GCAAAAGGATGCCTTCGACAA		T/C	Intronic
D6Int102	*Cascl*_Ex4	SNP	AAACCCAAGCAAACAGACCA GCAAAAGGATGCCTTCGACAA		A/G	Intronic
D6Int103	*Cascl*_Ex3	SNP	AAACATGGGCTGATCACCTG ATCCTAGTGCAGGCCTGTGT		G/A	Intronic
D6Int104	*Cascl*_Ex3	SNP	AAACATGGGCTGATCACCTG ATCCTAGTGCAGGCCTGTGT		G/T	5'UTR
D6Int105	*Cascl*_Ex3	SNP	AAACATGGGCTGATCACCTG ATCCTAGTGCAGGCCTGTGT		A/T	5'UTR
D6Int106	*Cascl*_Ex3	SNP	AAACATGGGCTGATCACCTG ATCCTAGTGCAGGCCTGTGT		G/A	5'UTR
D6Int107	*Cascl*_Ex3	SNP	AAACATGGGCTGATCACCTG ATCCTAGTGCAGGCCTGTGT		T/C	5'UTR
D6Int108	*Cascl*_Ex3	SNP	AAACATGGGCTGATCACCTG ATCCTAGTGCAGGCCTGTGT		T/C	5'UTR
D6Int109	*Cascl*_Ex3	SNP	AAACATGGGCTGATCACCTG ATCCTAGTGCAGGCCTGTGT		T/G	5'UTR
D6Int9	*Cascl*_Ex2	SSLP	GGGAACCAGGTCCTCTTCAA GACCAGAATGCTGGACATGG	33P		
D6Int110		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int111		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int112		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int113		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int114		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		

D6Int115		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int116		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int117		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int118		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int119		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int120		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int121		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int123		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int124		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int125		SEQ	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	sequence		
D6Int44	*Ghiso_Ex1*	SNP	TCGAGTCCCATACACTGCTG AGCTCTTGAGAAGGGCGAAT		A/T	5'UTR
D6Int126	*Ghiso_Ex1*	SNP	TCGAGTCCCATACACTGCTG AGCTCTTGAGAAGGGCGAAT		T/C	5'UTR
D6Int127	*Ghiso_Ex1*	SNP	TCGAGTCCCATACACTGCTG AGCTCTTGAGAAGGGCGAAT		T/A	5'UTR
D6Int128	*Ghiso_Ex1*	SNP	TCGAGTCCCATACACTGCTG AGCTCTTGAGAAGGGCGAAT		C/G	Intronic
D6Int129	*Ghiso_Ex1*	SSLP	TCGAGTCCCATACACTGCTG AGCTCTTGAGAAGGGCGAAT	32P	5 bp Ins	Intronic
D6Int130	*Ghiso_Ex1*	SNP	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT			
D6Int131	*Ghiso_Ex2*	SNP	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT			
D6Int132	*Ghiso_Ex2*	SSLP	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT	32P		
D6Int133	*Ghiso_Ex2*	SNP	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT			
D6Int134	*Ghiso_Ex2*	SNP	TCGAGTCCCATACACTGCTG AGCTCTTGAGAAGGGCGAAT		C/T	Intronic
D6Int135	*Ghiso_Ex2*	SNP	TCGAGTCCCATACACTGCTG AGCTCTTGAGAAGGGCGAAT			
D6Int136	*Ghiso_Ex2*	SNP	CAGAGCCAAAGTGGGAATGT AGCTCTTGAGAAGGGCGAAT			
D6Int137	*Ghiso_Ex2*	SNP	TCGAGTCCCATACACTGCTG AGCTCTTGAGAAGGGCGAAT		T/G	Intronic
D6Int42	*Ghiso_Ex3*	SNP	GGCCAATTCGTTACGAGGA CGATAGTGACAATCTCATCAGTAAATG		T/C	Intronic
D6Int10	*Ghiso_Ex3*	SNP	TTTTTAAAGCTGCTGTATCTTGGA CAAGGCCTCTAGCTCCTTCA		GGAGCAG[A/G]CTATTTT	D -> G
D6Int11	*Ghiso_Ex3*	SSLP	TGACTGCATAATTCTCAATAAATCACT TTCATCTCATTTTGTATTCCACTGA	33P		
D6Int138	*Kras2_4B*	SNP	TGTCCATCTACTCATCCAA GAGGGAGTGACCAAGATTA		G/A	3' UTR
D6Int139	*Kras2_4B*	SNP	TGTCCATCTACTCATCCAA GAGGGAGTGACCAAGATTA		C/A	3' UTR
D6Mit57	*Kras2* Ex 1	SSLP	TCTGATATCCAAGTCATCGTGG AAACCAAACAAAAGGAGTGGC	agar		
D6Int140	*Kras2_494*	SNP	*Kras2_494*			
Kras2_37	*Kras2* Ex 1	SSLP	GCCTCACCGAAAGACAGATCA AATCAAAGGAGCTGGGAACA	agar		
D6Int142	*Kras2* Ex 1	SNP	AAGCGCACGCAGACTGTAG GGCTGTTTAGATCAACAAGCTAAAT			
	Kras2 Ex 1					
D6Int143	*Kras2_Ex1g*	SNP	AAGCGCACGCAGACTGTAG GGCTGTTTAGATCAACAAGCTAAAT			
D6Int144	*Kras2_Ex1g*	SNP	AAGCGCACGCAGACTGTAG GGCTGTTTAGATCAACAAGCTAAAT			
D6Int145	*Kras2_Ex1g*	SNP	AAGCGCACGCAGACTGTAG GGCTGTTTAGATCAACAAGCTAAAT			
D6Int12	RS25_Intron	SSLP	TTCCAGACTTGCAAAGCAAA CTGAGGGAGCTCTGTGAAGG	33P		
D6Int43	SNP_157517_end_BAC	SNP	TCTCAATCTCACGCACCATC TTCCCCCCACACTGACTGTAGG		C/T	Genomic

D6Int146	SNP_157563_end_BAC	SNP	TCTCAATCTCACGCACCATC TTCCCCCACACTGACTGTAGG		T/C	Genomic
D6Int14	*Lmna-rs1*_Ex8	SSLP	ACCAAGTAAGTTCCAGGACAGC AGACAAATCATCTGCCTTGTGG	agar		Genomic
D6Int46	*Lmna-rs1*_Ex5	SNP	ACAGCCACCAACGGACTTT CATTGGGCTTGCACAAGAT		TGGGAG[A/G]ACAGAGC	E -> G
D6Int47	*Lmna-rs1*_Ex4	SNP	ATGTTGTCGGGGAACTGGTA AGATTCCAAACACGCACCTC		AATGGAC[A/G]CGCAGTC	H -> S
D6Int147	*Lmna-rs1*_Ex4	SNP	ATGTTGTCGGGGAACTGGTA AGATTCCAAACACGCACCTC			
D6Int148	*Lmna-rs1*_Ex4	SNP	ATGTTGTCGGGGAACTGGTA AGATTCCAAACACGCACCTC			
D6Int149	*Lmna-rs1*_Ex4	SNP	ATGTTGTCGGGGAACTGGTA AGATTCCAAACACGCACCTC			
D6Int56	*Lmna-rs1*_Ex2	SNP	TAGGGACTCACGGCAAGTTT GCACTCAGATGCCTTGCTTT		CxB)RI 5 Allele Balb	
D6Int153	*Lmna-rs1*_Ex1	SNP	AGCTCCTTCCATGACAGAGG CGAGGTCGTTGGTTGTCATA		C/T	
t48	*Lmna-rs1* 5' genomic	SSLP	ATCTCCAGGGAACTGCTTCA AGCATACCTGGGCATACAGC	32P		
D6Int15	*Lmna-rs1* 5' genomic	SSLP	TTGCATGCAGACACACACAA GTCAAATGGACACGTGGAAA	32P		
D6Int49		SSLP	CCTTGAGTTTGTGGCTCTCC AGGCCATGTGCAGTACCTTC	32P		
D6Int50		SSLP	GGGCAGCATAGCAAGTTTCA GCCAAGGTGAGGTTTTGAGA	32P		
D6Int51		SSLP	CCAGAACTGGTGCTGCTTTT ATCTCCAGGGAACTGCTTCA	32P		
D6Int52		SSLP	TCGTGTCAAGCCACTTAGTCA TTGCATGCAGACACACACAA	32P		
D6Mit14		SSLP	ATGCAGAAACATGAGTGGGG CACAAGGCCTGATGACCTCT	32P		
D6Int57		SSLP	AGGTTCTGCAAACTCCTCCA TGTCCCCTGAACCCATGTAT	32P		
D6Int45		SSLP	ACAGGCCCAGCTCAAGATTT ACCCAGGTCAGGTAGCCTCT	32P		
D6Int55	END YAC Y95M5E3	SNP	GGCTCCTGGTGGTTTGTCAA TTGCAACTCTCTGCTGGACT		A/G	
D6Int150	END YAC Y95M1F5	SNP	GGATGTGTGGTAATCGGAAGA CAAAGACACTTGCCACTAAGC		A/G	
D6Int151	END YAC Y88M6a6	SNP	GATCAACATCTGTCGGGAAGA TCTGGAGGTTGTGGTACCCA		T/G	
D6Int13		SSLP	CAAGCACACACACTTACAACCA TCATGGAGGCCTATAAGCAGA			

Table 4 159 SNPs/SSLPs identified in the nucleotide sequence analysis of genomic fragments for a total of 20000 bp in the *Pas1* region.

Susceptibility	Strain	Haplotype type	Reference
Susceptible	A/J	A/J	Manenti G., 2002 Malkinson A.M.,1983 Falconer D.S.,1962
	SWR/J	A/J	Manenti G.,1997 Malkinson A.M.,1983
	FVB/NJ	A/J	Shafarenko M.,1997 Mahler J.F., 1996
	STS/A	A/J	Van der valk M.A. 1981
Intermediate	CBA	A/J	Ryan J., 1987 Maltoni C., 1989
	BALB/CJ	A/J	Manenti G., 1997 Malkinson A.M.,1983 Trainin N., 1964
	129/SvJ	A/J	Malkinson A.M.,1983
	RF/J	A/J	Ryan J., 1987 Maltoni C., 1989
	PL/J	A/J	Malkinson A.M.,1989
	LP/J	A/J	Smith G.S., 1973
	ST/Bj	A/J	Malkinson A.M.,1989
	MA/MyJ	A/J	Hoag W.G., 1963
Resistant	P/J	C57BL/6	Hoag W.G., 1963
	SM/J	C57BL/6	Malkinson A.M.,1989
	AKR/J	C57BL/6	Malkinson A.M.,1983 Malkinson A.M.,1989
	C58/J	C57BL/6	Thaete L.G., 1991
	DBA/1J	C57BL/6	Trainin N., 1964
	DBA/2J	C57BL/6	Smith G.S., 1973 Malkinson A.M.,1983
	C57L/J	C57BL/6	Malkinson A.M.,1983
	C3H/HeJ	C57BL/6	Malkinson A.M.,1983 Hoag W.G., 1963
	C57BL/6J	C57BL/6	Malkinson A.M.,1983 Malkinson A.M.,1989
	C57BL/10J	C57BL/6	Malkinson A.M.,1983
	SJL/J	C57BL/6	Malkinson A.M.,1989
	C57BR/cdJ	C57BL/6	Malkinson A.M.,1983

Table 5 Inherited predisposition to lung tumorigenesis in 24 mouse inbred strains.

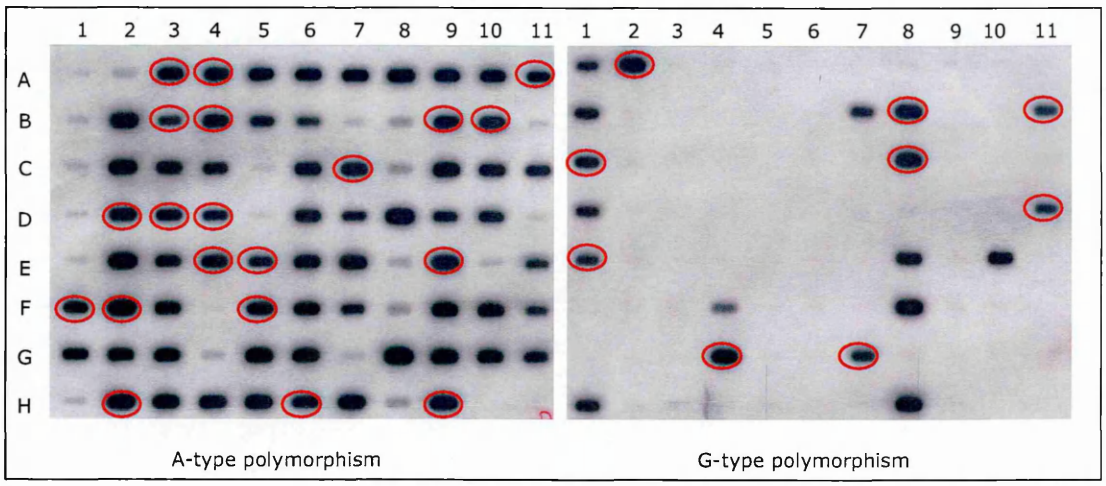

Figure 6 Example of detection of a SNP polymorphism (D6Int10) by allele-specific oligonucleotide (ASO) hybridization in 86 inbred strains. DNA samples were amplified by PCR, blotted on nylon membranes, and hybridized with either a radiolabeled A-type oligonucleotide (on left) or a radiolabeled G-type oligonucleotide (on right). The mouse strains studied in my thesis are in red. C1: 129/SvJ; E1: A/J; F1: AKR/J; A2: Balb/CJ; D2: C3H/HeJ; F2: C57BL/10J; H2 C57BL/6J; A3: C57L/J; B3: C57BR/cdJ; D3: C58/J; A4: CBA; B4: CE/J; D4: DBA1/J; E4: DBA2/J; G4: FVB/NJ; E5: LP7J; F5: MA/MyJ; H6: NZB/B1NJ; C7: P/J; G7: PL/J; B8: RF/J; C8: RIIIS/J; B9: SJL/J; E9: SM/J; H9: ST/BJ; B10: SWR/J; A11: NGP/N; B11: STS/A; D11: O20/A.

Subsequently, a Medline search was conducted to increase the number of strains. We decided to include five new strains of known susceptibility to lung cancer. In particular, NGP/N (Malkinson, 1989) and O20/A (Bentvelzen, 1966) are susceptible to lung tumorigenesis, RIIIS/J express intermediate susceptibility (Malkinson and Beer, 1983; Falconer and Bloom, 1962) while NZB/B1NJ (Malkinson and Beer, 1983) and CE/J (Festing and Blackmore, 1971) are resistant.

Position (kb)	*Marker*	*Gene*	*PCR primers*
142931483	*D6Int152*		CCTCTAGCCTCCTTGGGAAC
			TTCTGTCCTTTGGGGAAGG
145027489	*D6Int20*		TGCAGAGACTCACAATTCACG
			AGGCCCTGTGCTGGCTAAG
145175879	*D6Int16*		TGGGCCATGCTACAAGTTCT
			CCATGTTTGCTCTTGTGCTG
145230814	*D6Int21*		TGGAGTGCTCTGAGCTGTTG
			CCTATCTGTCAGTTTTATCCCTTGT
145337579	*D6Int17*		CAATGCGGGTTTATCTGTCA
			TACATGCACGACCCTCACAT
145346401	*D6Int24*		GGGAAATTGGTGGGAAGAAG
			CCTTGACTTTTGTTTCCTACCC
145363293	*D6Int25*		TGTCTAACTTCCCAGACCAACA
			CAAAGTGGCAGTGTCCTCAA
145389499	*D6Int26*		GGCATCTCATTCCTTCTTTCA
			CTGCTTGAATGCTTGGGTTT
145398672	*D6Int27*		TGCATTCGTGTGCACATTC
			GTGCCCTTACCCACTGTCAT
145403158	*D6Int18*		ATCCCACAACCTCATTCCAA
			CCTCCATTCTGTATGGCAAGA
145434169	*D6Int29*		CCCTGATACCACACATGCAC
			CTAGCTGGGCCAAGAAAGC
145506667	*D6Int3*	*Lrmp*	GCCATCTCAGTGGAGGCTTT
			CATGTCAGAGATGAGCTGTGC
145520113	*D6Int31*	*Lrmp*	AAACACGATATAGGCCCATTTAATA
			GTTGCTTTTCTTGGGCTGGT
145532679	*D6Int5*	*Lrmp*	TGAGCAGTCTCTCCAACCTCA
			TCTGCCAGTACCCACGTACA
145556176	*D6Int6*	*Lrmp*	TCCAAAATGGCCCTGCTTAT
			GTTGCCCTCTGACTTCCACA
145559317	*D6Int36*	*Lrmp*	GGGTGCTGTTGCTTAGTGGT
			ACACGTTTGTGGTGGTGTGA
145595636	*D6Int9*	*Las1*	GGGAACCAGGTCCTCTTCAA
			GACCAGAATGCTGGACATGG
145599886	*D6Int129*	Ghiso	TCGAGTCCCATACACTGCTG
			AGCTCTTGAGAAGGGCGAAT
145599956	*D6Int132*	Ghiso	TCGAGTCCCATACACTGCTG
			AGCTCTTGAGAAGGGCGAAT
145605994	*D6Int11*		TGACTGCATAATTCTCAATAAATCACT
			TTCATCTCATTTTGTATTCCACTGA
145634766	*Kras2_37*	*Kras2*	CAGGGCTGCATAGTAAGA
			GGGTGTTAGGGAACCATA
145644184	*D6Int12*		TTCCAGACTTGCAAAGCAAA
			CTGAGGGAGCTCTGTGAAGG
145788908	*D6Int14*		ACCAAGTAAGTTCCAGGACAGC
			AGACAAATCATCTGCCTTGTGG
145833516	*D6Int48*		ATCTCCAGGGAACTGCTTCA
			AGCATACCTGGGCATACAGC
145836112	*D6Int15*		TTGCATGCAGACACACACAA
			GTCAAATGGACACGTGGAAA
145931230	*D6Int52*		GATGTGACACAGCAGGATGC
			ACCATGGGATCCAGAGATCA
146035280	*D6Int45*		ACCCAGGTCAGGTAGCCTCT
			ACAGGCCCAGCTCAAGATTT

Table 6 Simple sequence length polymorphism with PCR primers used for the analysis of the *Pas1* locus

Position (kb)	Marker	*Gene*	*Exon*	*GenBank Acc. #*	*SNP polymorphism*[a]	*Amino-acid variation*[a]
145501932	*D6Int30*			AY490381	C145501932T	Genomic
145526703	*D6Int4*	*Lrmp*	2	AJ437328	G145526703A	G31D
145536150	*D6Int32a*	*Lrmp*	3	AY490382	A145536150G	D56G
145536157	*D6Int32b*	*Lrmp*	3	AY490382	G145536157C	L58F
145536163	*D6Int32c*	*Lrmp*	3	AY490382	C145536163T	S60S
145536417	*D6Int33*	*Lrmp*		AY490382	T145536417C	Intronic
145562248	*D6Int54*	*Lrmp*	17	AJ437329	G145562248A	G438R
145563172	*D6Int8*	*Lrmp*	18	AJ437330	T145563172C	L537P
145566018	*D6Int37a*	*Las1*		AY490383	G145566018T	Intronic
145580048	*D6Int38*	*Las1*	9	AY490384	C145580048T	Q278Q
145580256	*D6Int84*	*Las1*	9	AY490384	C145580256T	Intronic
145589112	*D6Int40*	*Las1*	5	AY490385	C145589112T	Intronic
145589248	*D6Int95*	*Las1*	5	AY490385	C145589248A	Intronic
145589273	*D6Int96*	*Las1*	5	AY490385	A145589273C	Intronic
145591455	*D6Int41*	*Las1*	4	AY490386	C145591455T	S60N
145591604	*D6Int101*	*Las1*	4	AY490386	T145591604C	Intronic
145591661	*D6Int102*	*Las1*	4	AY490386	A145591661G	Intronic
145599704	*D6Int44*	Ghiso	1	AY490387	A145599704T	5'-UTR
145599728	*D6Int126*	Ghiso	1	AY490387	T145599728C	5'-UTR
145599744	*D6Int127*	Ghiso	1	AY490387	T145599744A	5'-UTR
145599847	*D6Int128*	Ghiso		AY490387	C145599847G	Intronic
145599922	*D6Int130*	Ghiso		AY490387	G145599922A	Intronic
145599947	*D6Int131*	Ghiso		AY490387	G145599947A	Intronic
145600006	*D6Int134*	Ghiso		AY490387	C145600006T	Intronic
145600020	*D6Int135*	Ghiso		AY490387	G145600020T	Intronic
145600058	*D6Int137*	Ghiso		AY490387	T145600058G	Intronic
145603833	*D6Int42*	Ghiso		AY490388	T145603833C	Intronic
145603889	*D6Int10*	Ghiso	3	AJ437332	A145603889G	D28G
145635211	*D6Int142*	*Kras2*	1	AY490389	A145635211G	Y32Y
145635417	*D6Int143*	*Kras2*		AY490389	A145635417G	Intronic
145635480	*D6Int145*	*Kras2*		AY490389	A145635480G	Intronic
145656172	*D6Int43*			AY490390	C145656172T	Genomic
145656218	*D6Int146*			AY490390	T145656218C	Genomic
145812513	*D6Int46*	*Lmna-rs1*	5	AY490391	T145812513C	E258G
145814801	*D6Int47*	*Lmna-rs1*	4	AY490392	T145814801C	H218S
145824723	*D6Int56*	*Lmna-rs1*		AY490393	C145824723T	Intronic
145828658	*D6Int153*	*Lmna-rs1*		AY490394	C145828658T	Intronic

[a]Polymorphic markers are labelled such that the C57BL/6J allele is followed by the position in the Ensembl alignment and then the A/J allele.

Genomic and intronic polymorphisms occur in intergenic and intragenic regions, respectively.

Table 7 Single nucleotide polymorphisms used for the analysis of the *Pas1* locus

3.1.2 *Pas1* haplotype

The 468-kb region defined a core *Pas1* haplotype characterized by two main types: an A/J and a C57BL/6J type. The boundaries of the core were physically defined by the decay of its association to alleles at <145.346 Mb (*D6Int24*) and at >145.815 Mb (*D6Int47*) from the chromosome 6 centromere.

In this region we analysed 54 markers, 42 (78%) of which showed perfect segregation with the haplotype, that is, no strain of the A/J type carried a C57Bl/6J- type polymorphism and vice versa. The perfect segregant markers included both coding and non-coding polymorphisms; the latter were located either in the intragenic or intergenic regions of six genes mapping in the haplotype. (Tab. 8)

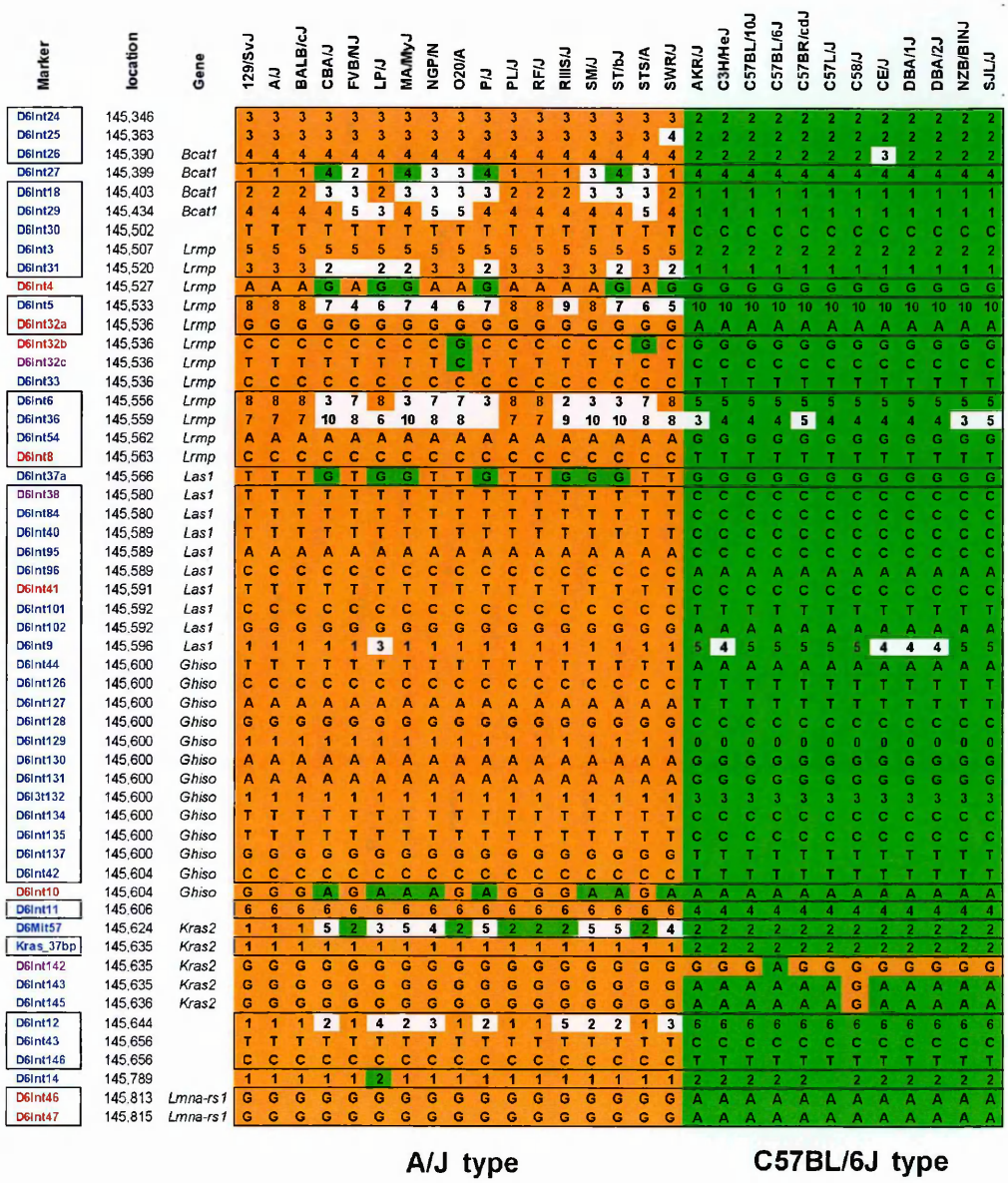

Table 8: Haplotype map of distal Chromosome 6 region including the *Pas1* locus. The mouse strains screened are indicated at the top of the figure, while the 54 genetic markers from the core *Pas1* haplotype region are listed on the left in colors to indicate molecular variations in non-coding regions (blue) or in coding regions (magenta, no amino acid change; red, amino acid change). Polymorphisms observed are shown in each cell and those present in A/J and in C57BL/6J are indicated in green and yellow, respectively. SNPs are shown as nucleotides according to the Ensembl database. SSLP-based polymorphisms are shown as allele numbers, whose values are inversely correlated with DNA fragment size. The location column indicates the physical position (in megabases) from the centromere according to the Ensembl database. The haplotype blocks and the markers that segregate appropriately according to the A/J- and C57BL/6J-type are boxed.

3.1.3 Coding region nucleotide polymorphism analyses

SNPs analyses revealed that all these genes, except *Bcat1,* showed nucleotides polymorphisms between A/J and C57BL/6J mice that produced amino acid changes (Tab.9).

Gene	Marker	Position	A/J allele	C57BL6/J allele	Amino-acid change	Exon	SNP type
Lrmp	D6Int4	145526703	A	G	G31D	2	Non-synonymous
Lrmp	D6Int32a	145536150	G	A	D56G	3	Non-synonymous
Lrmp	D6Int32b	145536157	C	G	L58F	3	Non-synonymous
Lrmp	D6Int32c	145536163	T	C	S60S	3	Synonymous
Lrmp	D6Int54	145562248	A	G	G438R	17	Synonymous
Lrmp	D6Int8	145563172	C	T	L537P	18	Non-synonymous
Casc1	D6Int41	145591455	A	G	S60N	4	Non-synonymous
Casc1	D6Int38	145580048	A	G	Q278Q	9	Synonymous
Ghiso	D6Int10	145603889	G	A	D28G	3	Non-synonymous
Kras2	D6Int142	145635211	G	A	No	1	Synonymous
Lmna-rs1	D6Int47	145814801	A	G	H218S	6	Non-synonymous
Lmna-rs1	D6int46	145812513	A	G	E258G	7	Non-synonymous

Table 9 Nucleotide polymorphism between A/J and C57BL/6J mice in *Lrmp*, *Casc1*, *Ghiso*, *Kras2* and *Lmna-rs* gene. Coding polymorphism linked with the *Pas1* haplotype are shown in red.

Lrmp gene carries two amino acid-changing polymorphism tingly linked with the *Pas1* haplotype: a D56G polymorphism (*D6Int32a*) in exon 3 and a L537P polymorphism (*D6Int8*) in exon 18. In particular the D6Int8 polymorphism segregates with both the *Pas1* haplotype and the susceptibility inflammatory response in mouse lines phenotypically selected for high and low acute inflammatory response (Maria *et al.*, 2003). Additional amino acid changes (*D6Int4* and *D6Int32b*) were found, but they are not linked with the

Pas1 haplotype while the D6Int32c polymorphism in exon 3 is a silent polymorphism. (Fig.7)

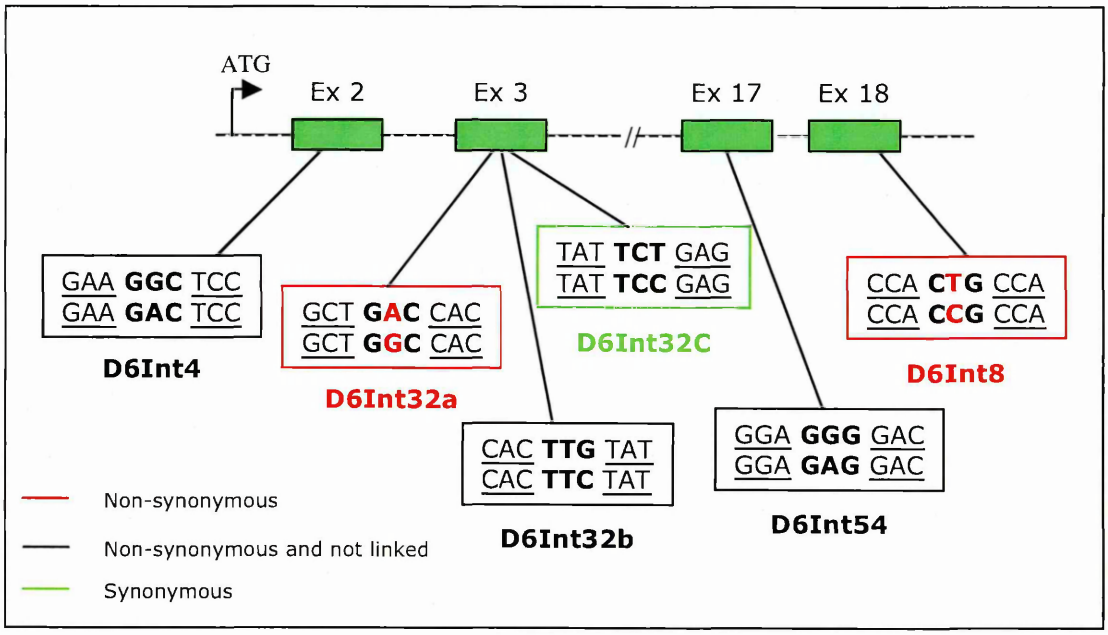

Figure 7 Amino-acid changed polymorphisms in *Lrmp*. The boxes indicate the A/J (down) and C57BL6/J (on the top) sequence.

Analysis of *Casc1* showed a S60N polymorphism (*D6Int41*) in exon 4 and silent Q278Q polymorphism (*D6Int38*) in exon 9. (Fig.8)

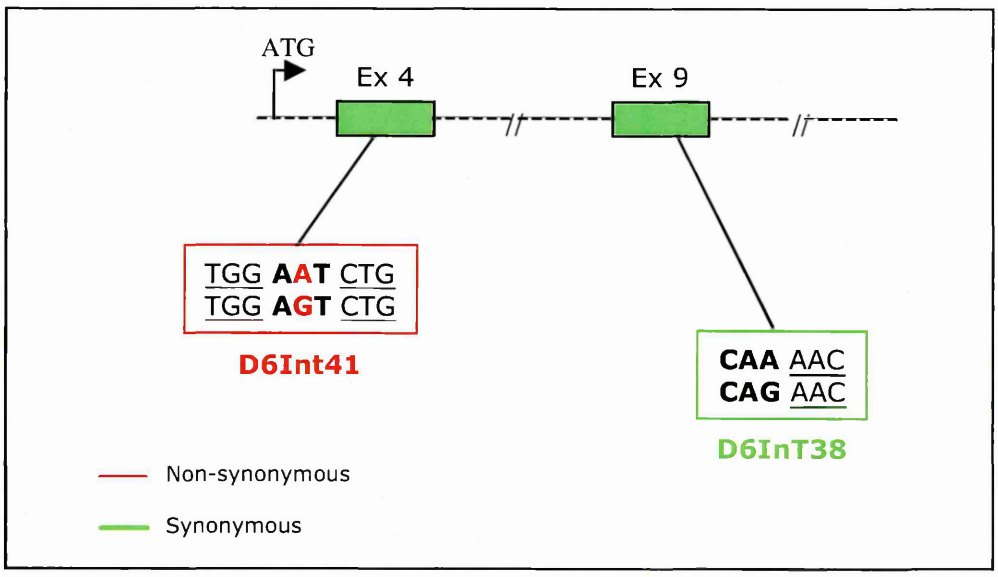

Figure 8 Amino-acid changing polymorphisms in *Casc1*. The boxes indicate the A/J (down) and C57BL6/J (on the top) sequence.

We found a polymorphism non-linked with the haplotype causing an amino acid change on exon 3 of *Ghiso* gene (*D6Int10*) (Fig.9) and a silent polymorphism (*D6Int142*) on exon 1 in *Kras2* gene (Fig.10).

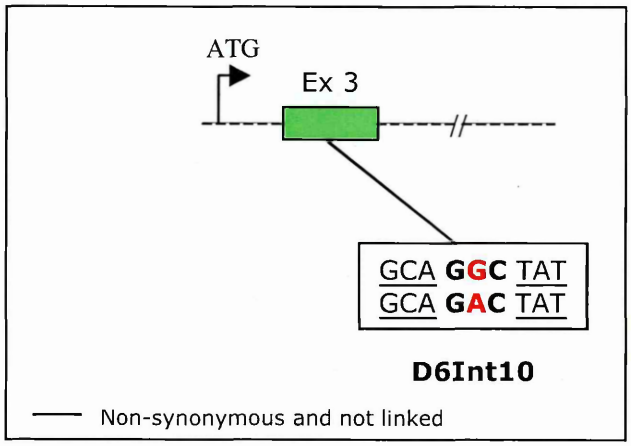

Figure 9 Amino-acid changing polymorphism in *Ghiso1*. The boxes indicate the A/J (down) and C57BL6/J (on the top) sequence.

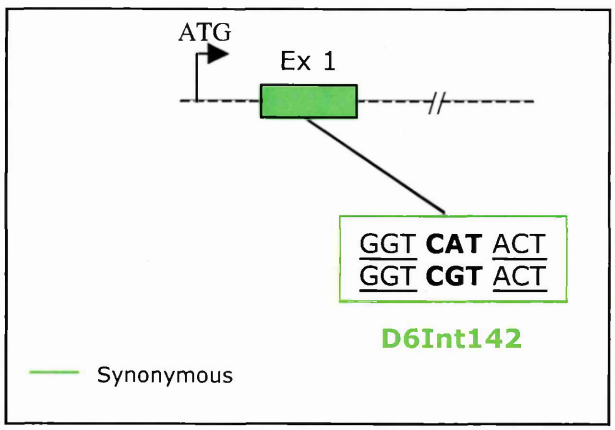

Figure 10 Polymorphism in *Kras2*. The boxes indicate the A/J (down) and C57BL6/J (on the top) sequence.

Lmna-rs1 carries a E258G polymorphism (*D6Int46*) in exon 5 and a H218S polymorphism (*D6Int47*) in exon 4. (Fig.11)

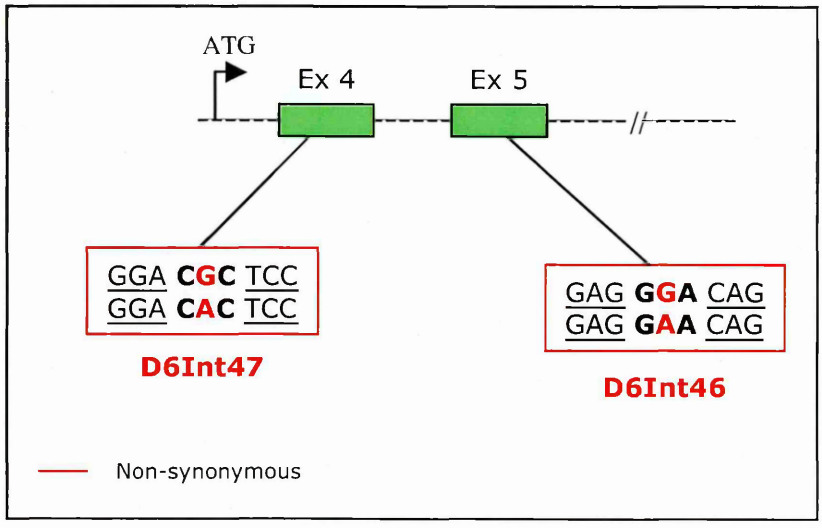

Figure 11 Amino-acid polymorphism in *Lmna-rs1*. The boxes indicate the A/J (down) and C57BL6/J (on the top) sequence.

3.2 CANDIDATE GENES FUNCTIONAL ANALYSIS

I continued my project by testing functional activity of candidate genes in vitro in order to study possible effects of mouse genes in human cells as a step toward the application of findings in mice to humans.

3.2.1 Allele-specific expression of candidate genes in mouse lung tumors

We carried out the expression analysis of the six candidate genes in four normal lungs of A/J and C57BL/6J mice, in five non-affected surrounding lung parenchyma of AB6F1 mice, and in nine urethane-induced lung tumors (3-5 mm in diameter) of AB6F1 hybrids.

We performed semiquantitative RT-PCR and we found that all the six genes are expressed at a similar level in the normal mouse lung of A/J, C57BL/6J, and AB6F1 mice while we found some differences in the analysis of tumor samples. This analysis revealed that transcript levels were similar for *Bcat1*, *Casc1*, *Kras2*, and *Ghiso* genes, while *Lrmp* showed a slightly reduced expression (about twofold) in six out of nine tumors. A difference in the expression was found in lung tumors where all genes but *Lmna-rs1* were expressed (Fig.12)
This result suggests that *Lmna-rs1* may represent a putative new tumor suppressor gene or a cell-type differentiation marker, whose expression is restricted to specific lung cell types that do not give rise to lung tumors.

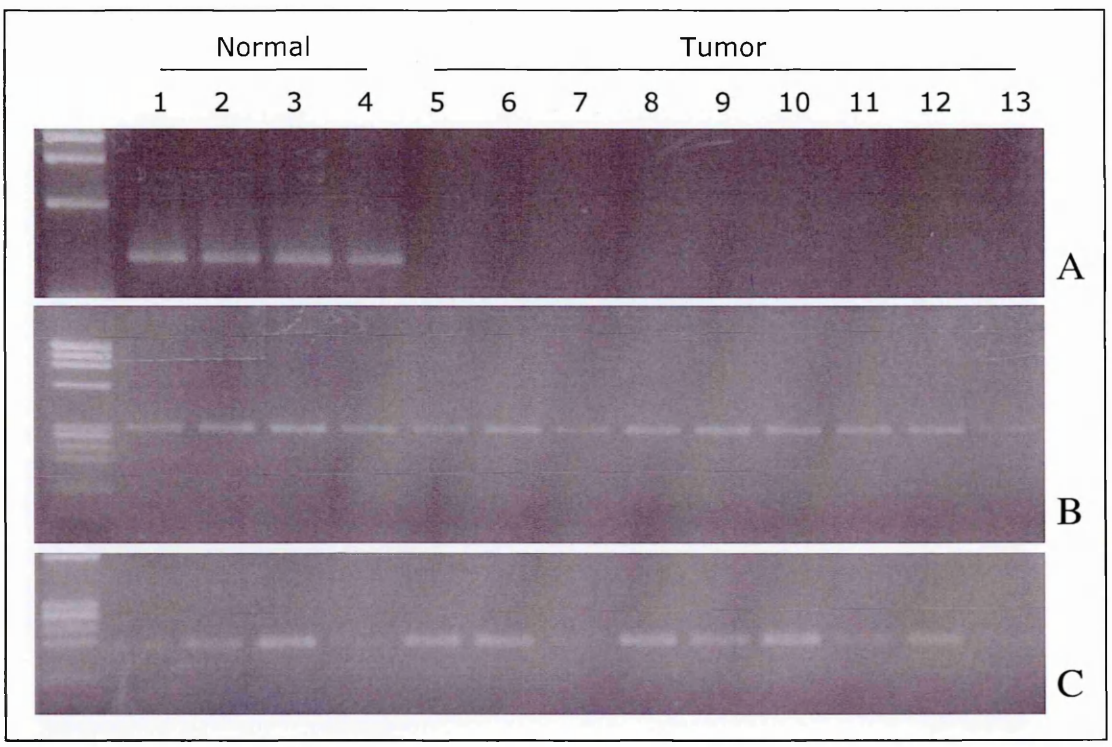

Figure 12 mRNA expression of the *Lmna-rs1* (A) *Ghiso* (B) and Gapdh (C) genes in mouse normal lung and lung tumors by RT-PCR. Normal lungs derived from A/J (lanes 1, 2) and C57BL/6J (lanes 3, 4) mice. Lung tumors derived from AB6F1 mice (lanes 5-13).

Allele-specific expression patterns of *Lrmp*, *Ghiso*, *Casc1*, and *Kras2* genes were analyzed in normal lungs and tumors from AB6F1

hybrids. *Lrmp* and *Ghiso* revealed no allelic differences in expression in normal lungs or tumors of these mice. Expression of both *Casc1* alleles was detected in the normal lungs, but mono-allelic expression of the 60N allele deriving from the A/J lung tumor-susceptible parent was found in all tumor samples analyzed for the S60N polymorphism in exon 4 (*D6Int41*) (Fig.13), suggesting allele-specific silencing/activation of the *Casc1* gene in lung tumors.

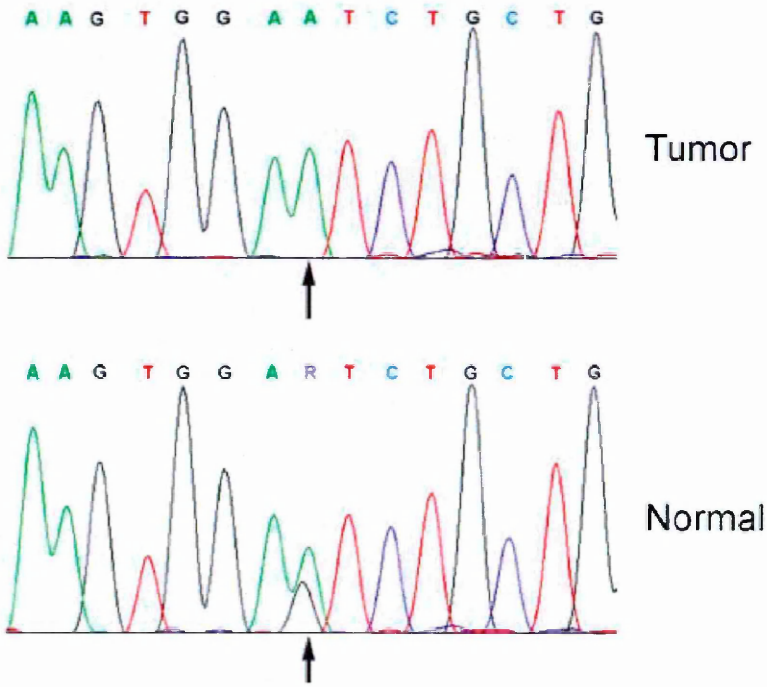

Figure 13 Representative nucleotide sequence pherograms of the *Casc1* gene by direct sequencing of RT-PCR products from normal lung and lung tumors of AB6F1 mice. Arrows indicate the mRNA expression of both A/J (green, A nt, 60N) and C57BL/6J (black, G nt, 60S) alleles in normal lung, whereas only the A/J allele is expressed in the lung tumors.

Analysis of *Kras2* allele expression using the herein identified *D6Int142* SNP (A145635211G, Tyr32Tyr) (Acc. # AY490389) polymorphism located in exon 1 of this gene revealed a slightly preferential expression of the A/J as compared to the C57BL/6J allele (2- to 3-fold ratio) in the 9 tumor samples (not shown). All tumors carried a codon 61 *Kras2* mutation (CAA→CTA in 5/9 tumors and CAA→CGA in the remaining 4 tumors), but the mutated allele

was expressed at heterozygosity in all tumors, with about 1:1 ratio of the wild-type to the mutated allele (Fig.14).

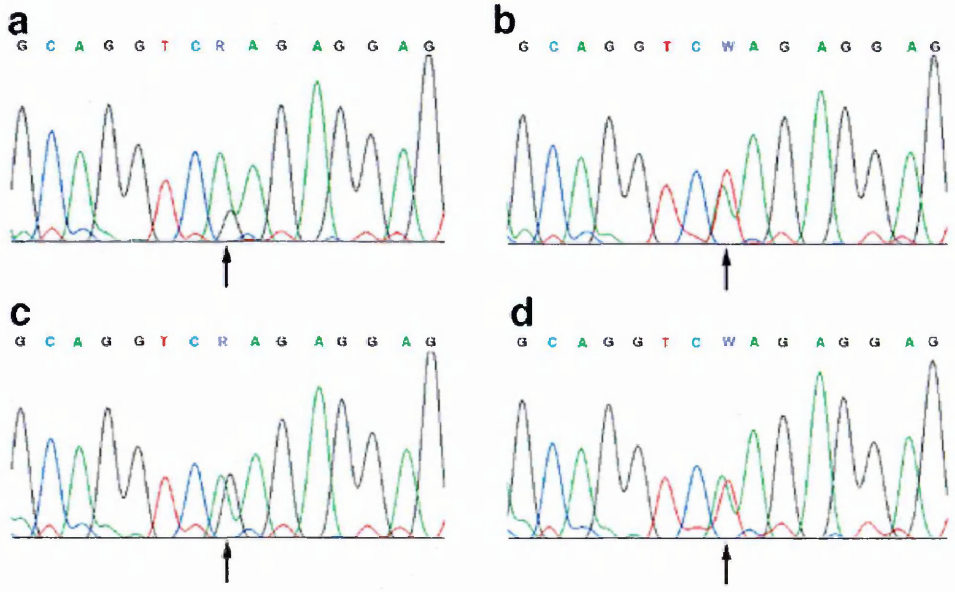

Figure 14 Representative nucleotide sequence pherograms of the *Kras2* gene by direct sequencing of RT-PCR products from urethane-induced lung tumor of AB6F1 mice, showing the mRNA expression of both the wild-type codon 61 (CAA), codon 61 CTA mutation (panels B and D), and codon 61 CGA mutation (panels A and C). Arrows indicate the position of the mutated nucleotide. Similar levels of wild-type and mutated codon 61 mRNA expression were detected in tumors C and D, whereas the wild-type allele was more highly expressed in sample A, and sample B showed slightly higher levels of the mutated allele.

3.2.1.1 Functional activity of candidate genes in vitro: clonogenic assay

To evaluate the effects of *Pas1* candidates on cell proliferation, we carried out a series of transfection experiments to determine whether these genes could promote or inhibit growth of two different human lung cell lines.

We initially tested the ability of the pEF6V5-His-TOPO vector to express recombinant candidate *Pas1* genes in human cells. We

transfected the two alleles of the candidate genes in HEK-293T cells, that is a human epithelial cell line, which expressed high levels of transfected proteins. (Thomas and Smart, 2005 in Galbiati *et al.*, 2005). HEK-293T cells were transiently transfected with 5μg recombinant plasmid DNA. After 24, 48 and 72h we pelleted the cells and then we extracted RNA. Candidate *Pas1* genes were RT-PCR-amplified with the primers used for subcloning. Analysis of mRNA expression revealed that all genes are well expressed and each allele is expressed at the same level at 24h, 48h and 72h after transfection. (Fig.15)

Then we performed a Western blot analysis by transiently transfecting HEK-293T cells to confirm these data and to look for the corresponding protein expression.
Western blot analysis confirmed that *Lmna-rs1*, *Ghiso*, *Lrmp* and *Casc1* gene are well expressed 24, 48 and 72h after transfection and allowed us to find the corresponding protein expression of all genes.

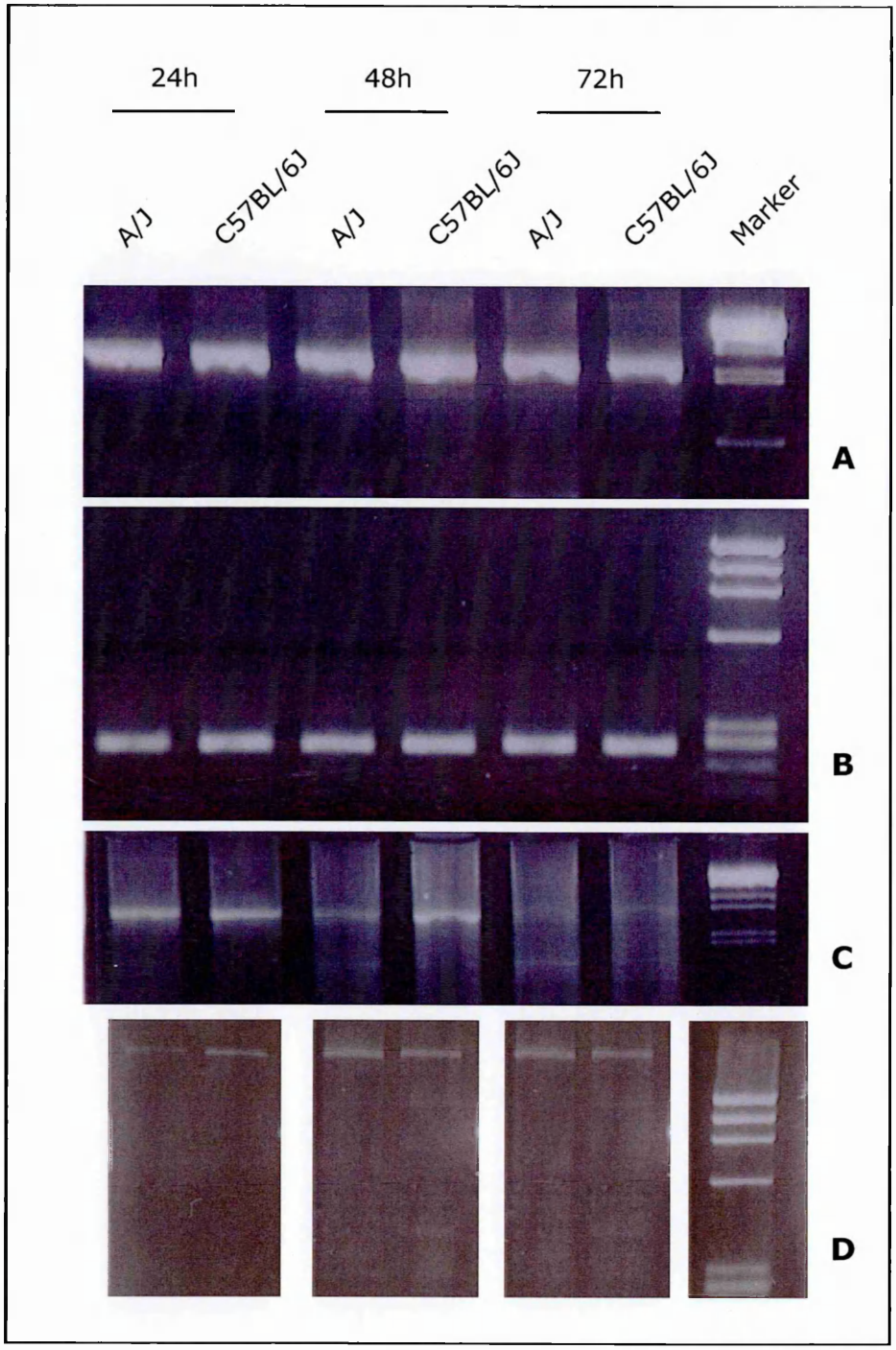

Figure 15 mRNA expression of *Lmna-rs1* (A), *Ghiso* (B), *Lrmp* (C) and *Casc1* (D) gene in HEK-293T cells. All genes are well expressed and each allele (A/J and C57BL/6J) is expressed at the same level 24-48 and 72h after transfection.

Once we demonstrated the capability of recombinant pEF6/V5-His-TOPO expression vector to express *Pas1* candidate proteins in human cells, we set up a stable transfection of two human lung

cancer cell lines (A549 and NCI-H520 cells) with the same recombinant vectors for the clonogenicity assay analysis.
A549 and NCI-H520 cells seeded in 6-multi-well plates at $3x10^5$ cells/well were transfected with 2µg recombinant plasmid DNA and 4 ùµl SuperFect (Qiagen) according to the manufacturer's protocol.
Cells transfected with the empty vector were used as a control. Two days after transfection, the appropriate concentration of antibody was added and transfectant clones were selected after 17 days for A549 and 26 days for NCI-H520 cells. Blasticidin-resistant clones were either fixed with methanol (100%) and stained with 10% Giemsa. We repeated at least three independent experiments.

The results obtained after the transfection of the two different cell lines were the same but the NCI-H520 cell line produces fewer and smaller clones than A549. The results were confirmed in all the experiments that we performed. The discrepancy in the size of clones is probably due to the different cells size and to the different cell grow rate.
In both A549 and NCI-H520 transfected cell lines, A/J and C57BL6/J alleles of candidate genes produced a variable number of clones. The number of clones of A/J and C57BL6/J alleles observed in *Lrmp* and *Ghiso* was the same and was similar to the vector alone (used as a control). While the A/J allele of *Lmna-rs1* produced more colonies than the C57BL6/J allele and the control (A and B in Fig.16), the A/J allele of *Casc1* and the empty vector produced much fewer colonies than the C57BL6/J allele did (C and D in Fig.16).

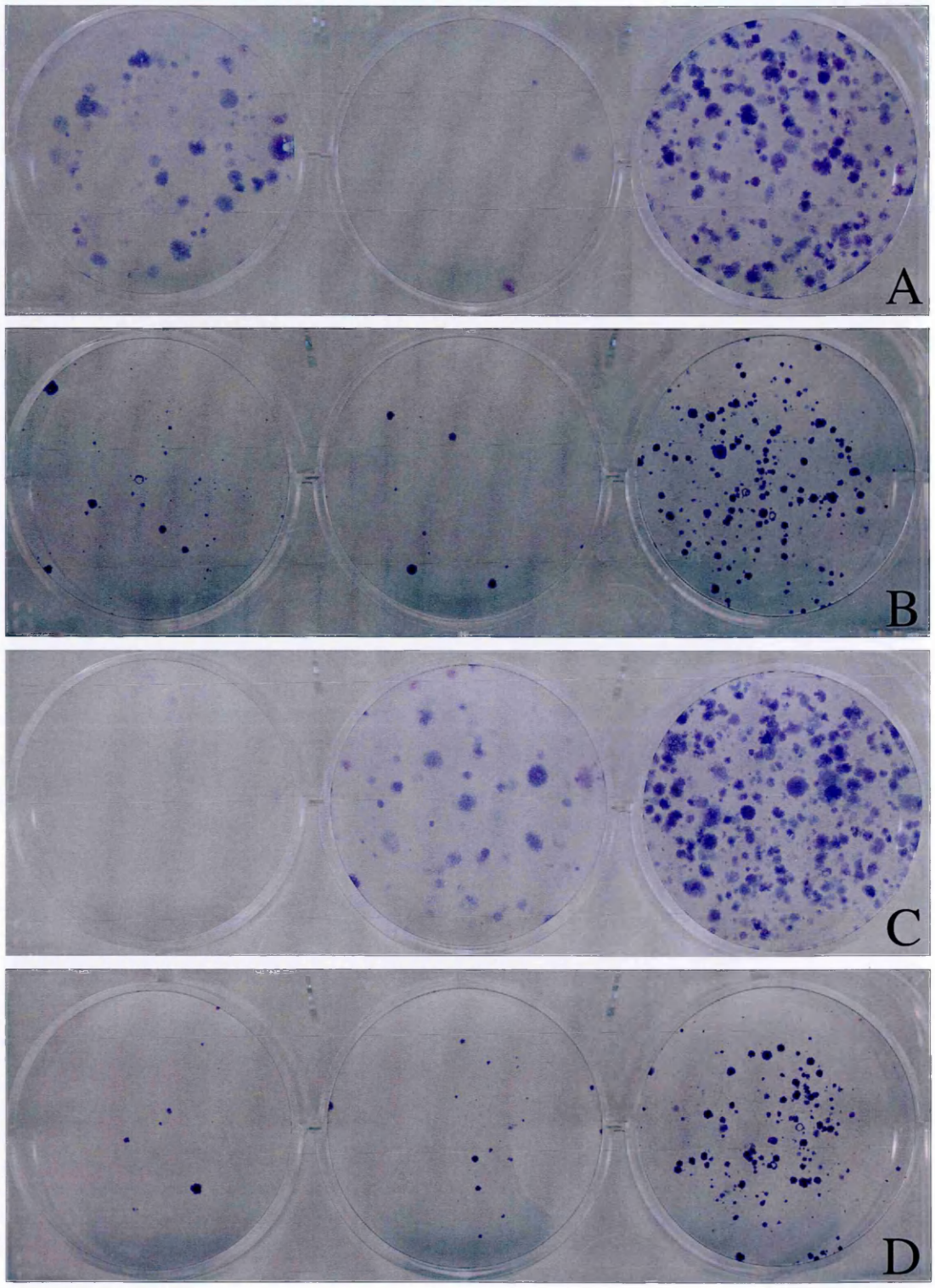

Figure 16 Qualitative analysis of colony formation in human lung cancer cell transfectants. Giemsa-stained plates reveal in vitro allele-specific colony formation activities of mouse *Lmna-rs1* (A and B) or *Casc1* (C and D) in human lung carcinoma A549 (A and C) or NCI-H520 (B and D) cells transfected with these genes. Colony formation in both cell types was inhibited for the C57BL/6J allele of *Lmna-rs1* and for the A/J allele of *Casc1*.

During the count, we decided to exclude colonies smaller than 0.5 mm in diameter. The evaluation of the number of stained colonies was affected by this decision. Therefore, for a more quantitative evaluation, we decided to perform an analysis based on the count of the total number of blasticidin-resistant cells in each well.

A549 cells stably transfected with the A/J allele of *Lmna-rs1* produced 34.6±2.8 (mean±SE) $x10^4$ cells/well *versus* 9.0±0.4 $x10^4$ cells/well produced by the C57BL/6 allele ($P<0.0001$). The same pattern was confirmed in NCI-H520 cells, with transfection of the A/J and the C57BL/6J allele producing 32.9±6.1 and 14.9±3.2 $x10^4$ cells/well, respectively ($P=0.020$). A549 cells transfected with the *Casc1* A/J allele produced 4.2±0.7 *versus* 25.9±4.2 $x10^4$ cells/well with the *Casc1* C57BL/6 allele, i.e., a 6-fold difference ($P<0.0001$). In NCI-H520 cells, transfection of the *Casc1* A/J allele produced 5.3±1.1 *versus* 26.6±3.0 $x10^4$ cells/well produced by the C57BL/6 allele, i.e., a 5-fold difference ($P<0.0001$).

No allele-specific effects were detected for *Ghiso* and *Lrmp* by cell count, in agreement with the qualitative observations of Giemsa-stained colonies. We found no differences between the two alleles when we transfected A549 and NCI-H520 cells. In particular, A549 cells transfected with *Lrmp* alleles produced somewhat fewer (~1.4-fold) numbers of cells/well ($P<0.05$), whereas *Ghiso* transfection showed no significant effects; in NCI-H520 cells, transfection with *Lrmp* led to a 4-fold reduction in the number of cells/well as compared to the empty vector transfection ($P<0.0001$). *Ghiso* transfection led to a 3-fold reduction in the number of cells/well as compared to the empty vector ($P>0.0001$). (Fig.17)

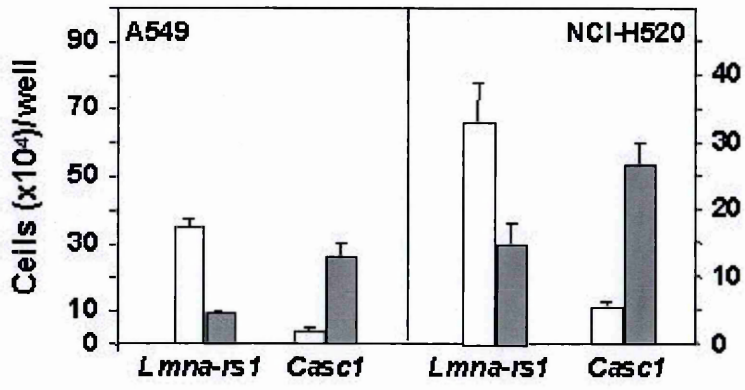

Figure 17 Quantitative analysis of allele-specific colony formation in human lung carcinoma A549 (left) or NCI-H520 (right) cells. A/J and C57BL/6J alleles showed statistically significant differences. Transfection with the empty vector produced 66 x 10^4 and 128 x 10^4 A549 and NCI-H520 cells, respectively.

3.2.2 Apoptosis analysis in candidate genes

We hypothesized that inhibition of clonogenicity by allelic forms of *Casc1* and *Lmna-rs1* genes was mediated by induction of apoptosis. Therefore, we tested whether PARP [Poly (ADP-ribose) polymerase] cleavage might underlie the allele-specific modulation of clonogenicity by candidate genes. PARP is a 116 kDa nuclear chromatin-associated enzyme and is a target of the caspase protease activity associated with apoptosis. During apoptosis PARP is cleaved from its 116 kDa intact form into 85 kDa and 25 kDa fragments. Apoptosis was analyzed in HEK-293T, A549, and NCI-H520 cells transiently transfected with A/J and C57BL/6J alleles of *Casc1*, *Lrmp*, *Ghiso* and *Lmna-rs1* genes. After 24, 48 and 72 h, cells were lysed and a western blot experiment was performed. We used as a positive control anti-Fas (CD95)-treated ovarian cancer cells (OAW42)), and as a negative control OAW42 not treated as well as non transfected HEK-293T, A549, and NCI-H520 cell lines.
We expected the presence of the two fragments in A/J allele of *Casc1* and in C57BL6/J allele in *Lmna-rs1*.
Western blot analysis revealed no evidence that PARP was cleaved into the 85 kDa fragment at 48 or 72h after transfection of any of the candidate genes allelic variants. (Fig.18)

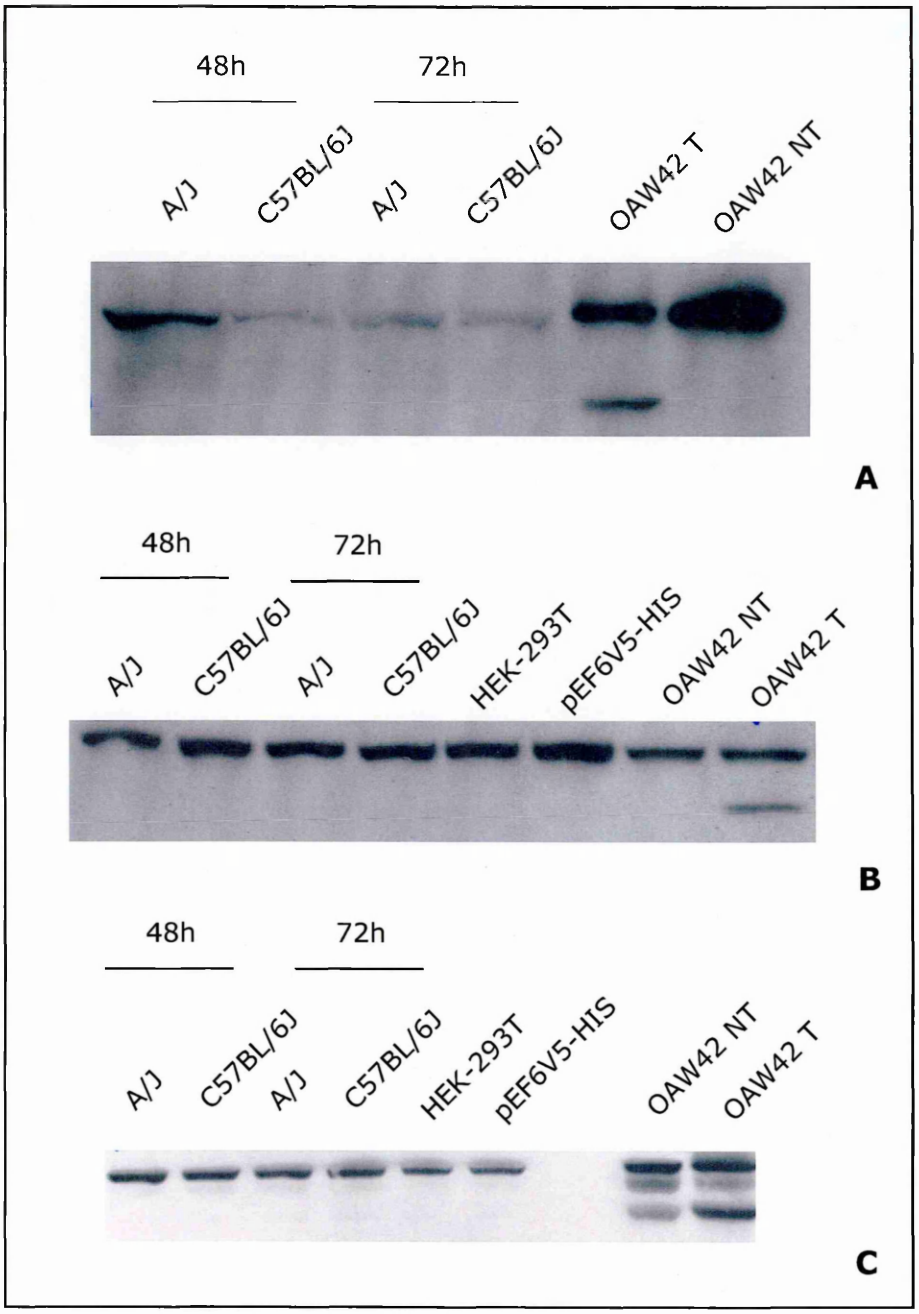

Figure 18 Apoptosis analysis in *Lrmp* (A), *Lmna-rs1* (B) and *Casc1* (C) by Western blotting in HEK-293T cell line for *Lrmp* and in A549 cell line for *Lmna-rs1* and for *Casc1*. Wild type cells were used as a negative control while anti-Fas (CD95)-treated ovarium cancer cells (OAW42) and not treated were used as a positive control. Our results revealed no evidence that PARP was cleaved at 48 and 72h after transfection of any of the candidate genes allelic variants.

3.2.3 Immunofluorescence analysis

The immunofluorescence analysis was performed in HEK-293T cells transiently transfected with calcium phosphate 24h after transfection. The immunofluorescence analysis revealed all recombinant *Pas1* candidate proteins localized mainly in the cytoplasm.

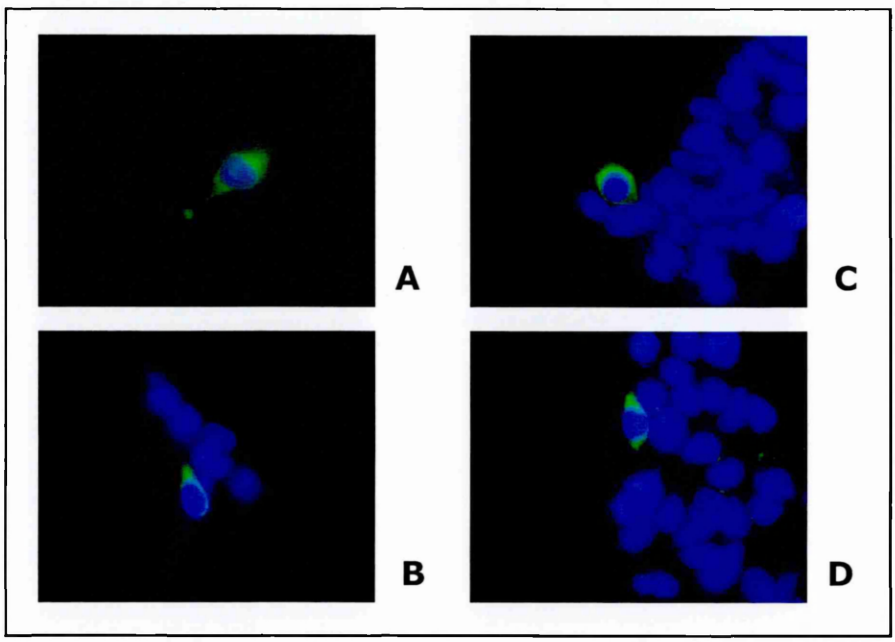

Figure 19 Immunofluorescence staining of protein encoded by *Lrmp* candidate gene in HEK-293T cell line. Cells were transfected with A/J (A and C) and C57BL6/J (B and D) full length cDNAs, grown for 48h (A and B) and 72h (C and D) after transfection, fixed and treated with anti-V5 monoclonal antibody followed by secondary antibody-Alexa Flour 488 conjiugate. (Nikon Eclipse E1000 fluorescence microscope. Magnification: x40).

Lrmp (Fig.19) and *Ghiso* (Fig.20) proteins were expressed at a very low level and in particular, *Ghiso* proteins were localized in small vacuoli around the nucleus.

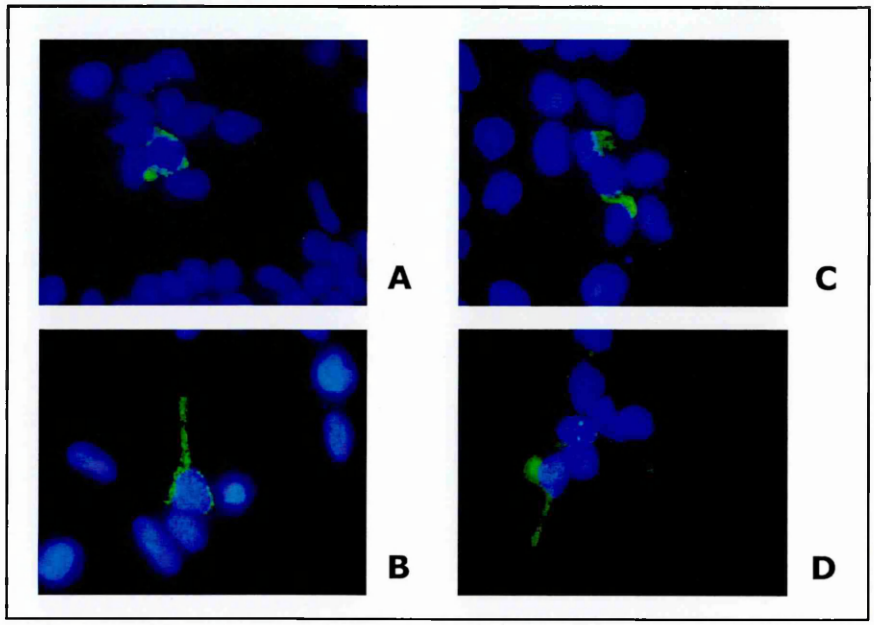

Figure 20 Immunofluorescence staining of protein encoded by *Ghiso* candidate gene in HEK-293T cell line. Cells were transfected with A/J (A and C) and C57BL6/J (B and D) full length cDNAs, grown for 48h (A and B) and 72h (C and D) after transfection, fixed and treated with anti-V5 monoclonal antibody followed by secondary antibody-Alexa Flour 488 conjiugate. (Nikon Eclipse E1000 fluorescence microscope. Magnification: x40).

Casc1 and *Lmna-rs1* proteins were expressed at a relatively high level. The *Casc1* (Fig.21) staining in the cytoplasm suggests that the protein is probably associated with the cytoskeleton fibers. While *Lmna-rs1* staining was uniformly diffuse (Fig.22)

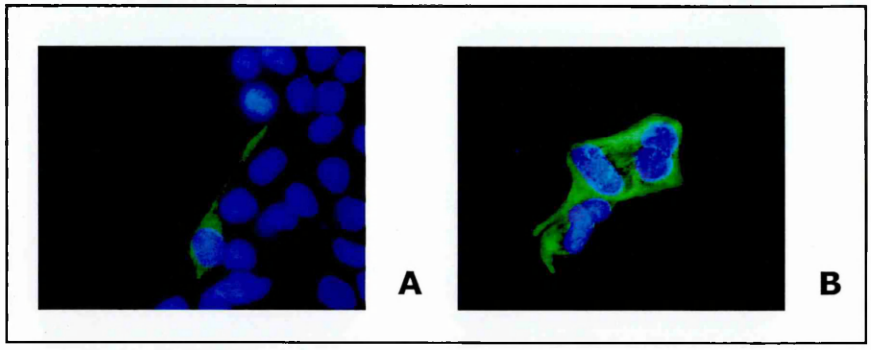

Figure 21 Immunofluorescence staining of protein encoded by *Casc1* gene in HEK-293T cell line. Cells were transfected with A/J (A) and C57BL6/J (B) full length cDNAs, grown for 48h (A and B) and 72h (not shown) after transfection, fixed and treated with anti-V5 monoclonal antibody followed by secondary antibody-Alexa Flour 488 conjiugate. (Nikon Eclipse E1000 fluorescence microscope. Magnification: x40).

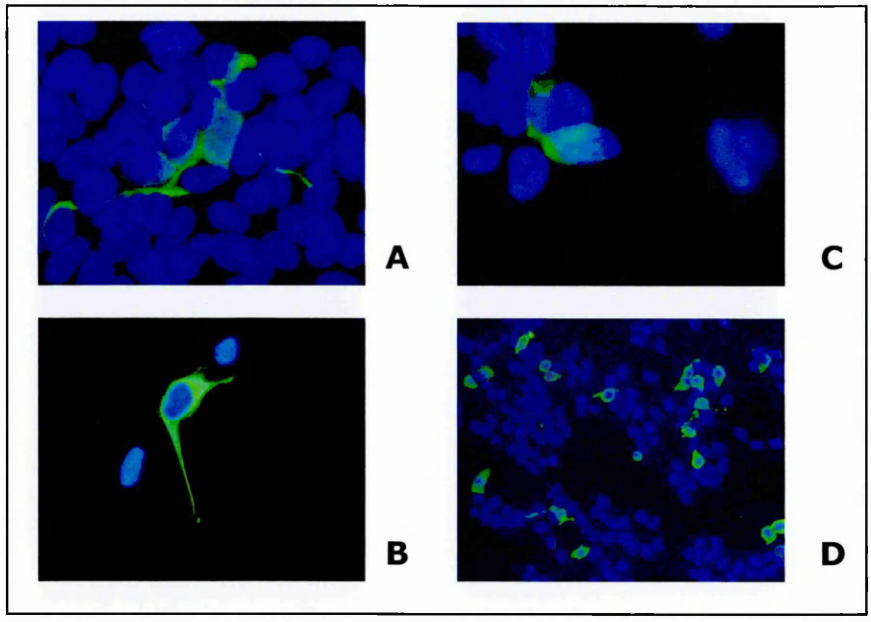

Figure 22 Immunofluorescence staining of protein encoded by *Lmna-rs1* candidate gene in HEK-293T cell line Cells were transfected with A/J (A and C) and C57BL6/J (B and D) full length cDNAs, grown for 48h (A and B) and 72h (C and D) after transfection, fixed and treated with anti-V5 monoclonal antibody followed by secondary antibody-Alexa Flour 488 conjiugate. (Nikon Eclipse E1000 fluorescence microscope. Magnification: x40 for A, B and C. Magnification: x20 for D.

Chapter 4 DISCUSSION AND CONCLUSION

In the present study, we identified and genotyped 65 genetic markers in the *Pas1* locus region around *Kras2* gene in a population of 29 laboratory mouse inbred strains with known susceptibility to lung carcinogenesis. This allowed us to construct a high-density *Pas1* haplotype map. The boundaries of the *Pas1* core haplotype were physically defined by the decay of its association to alleles at <145.346 Mb (*D6Int24*) and at >145.815 Mb (*D6Int47*) from the Chromosome 6 centromere. The distal boundary of the *Pas1* locus was also defined by a genetic recombination, which was nearly coincident with the haplotype boundary, in the CXB5 mouse line.

A/J- and C57BL/6J-type genetic variations segregated almost perfectly over the 468-kb haplotype region in the 29 laboratory strains. A recent study used A/J x C57BL/6 advanced intercross lines (AIL) to shorten the *Pas1* locus interval, but it provided no advantage over LD analysis, since it suggested placement of the *Pas1* locus in a 1.3-Mb interval that includes the previous 0.45-Mb LD region and ~27 known or predicted genes (Wang et al., 2003). Analysis of A/J x C57BL/6 congenic strains narrowed the *Pas1* locus interval to a 0.5-Mb region containing 12 putative genes (Zhang et al., 2003). Six of these 12 genes, chosen on the basis of mRNA expression in lungs, have been analyzed for coding polymorphisms and in vitro allelic effects; five of the six genes showed no allelic-specific effects in vitro and their candidacy was excluded. However, that study cannot be considered conclusive since these five expressed genes might also modulate lung tumorigenesis by mechanisms other than those affecting tumor growth (Hanahan & Weinberg, 2000).

We directed our attention to six putative *Pas1* candidate genes that map in this 468-kb region. The biological or biochemical functions of most of these genes are unknown:

Bcat1: maps to location 145.845604-145.895256. It has 11 exons and the transcript length is 2,508 bps.

This gene encodes the cytosolic form of the enzyme branched-chain amino acid transaminase and catalyzes the first step of branched-chain amino acid catabolism. This enzyme catalyzes the reversible transamination of branched-chain alpha-keto acids to branched-chain L-amino acids essential for cell growth. Two different clinical disorders have been attributed to a defect of branched-chain amino acid transamination: hypervalinemia and hyperleucine-isoleucinemia. As there is also a gene encoding a mitochondrial form of this enzyme, mutations in either gene may contribute to these disorders. Furthermore, *Bcat1* appears to be involved in apoptosis (Eden and Benvenisty, 1999).

Lrmp: maps to location 145.968258-146.021372. It has 19 exons and the transcript length is 2,238 bps.

The lymphoid-restricted membrane protein encoded by this gene has 539 residues and it is expressed in a developmentally regulated manner in lymphoid cell lines and tissues. The protein is localized to the cytoplasmic face of the endoplasmic reticulum and has a coiled-coil domain in the middle-third of the protein and a carboxyl-terminal membrane anchor (Behrens et al 1994).

Lrmp expression was originally described as lymphoid cell-specific, with highest expression levels in pre-T, pre-B, and mature B cells (Behrens et al., 1994). However, we detected significant *Lrmp* expression levels in lung tumors (not shown); clarification of whether *Lrmp* expression occurs in normal epithelial lung cells or is instead restricted to lung-associated lymphoid cells awaits immunohistochemical analysis of normal lung. The Jaw1-homologous protein Mrvil,

mapping to Chromosome 7, causes leukemia in mice upon activation by retroviral integration (Shaughnessy, Jr. et al., 1999), suggesting an important role for this class of integral membrane proteins in cancer and, possibly, for *Lrmp* polymorphisms in lung cancer susceptibility. The role of lymphoid cells and, more widely, of the immune system in lung tumorigenesis is suggested by enhancement of lung adenoma formation in neonatally thymectomized mice treated with 7,12-dimethylbenz(a)anthracene or urethane (Trainin et al., 1967). Moreover, splenectomy inhibited DMBA-induced lung tumor incidence in mice (Akamatsu, 1975), and inflammation is a known modulator of tumor development, both in mice and humans (Brownson & Alavanja, 2000; Mayne et al., 1999; Askling et al., 1999; Bauer et al., 2001).

The role of the *Lrmp* gene in lung tumorigenesis is not known; this gene is involved in antigen processing (Snyder et al., 1997).

Casc1: this gene corresponds to the recently identified gene (Acc. # AY423542) (Zhang et al., 2003). Our *Casc1* transcript (Acc. # AY485607, AY485608), compared to the AY423542 transcript, is 70 bp longer in the 5′-UTR, most likely represents the full-length transcript of the gene. We identified and then cloned *Casc1* using a different program to identify coding regions in the genome (NIX analysis package). *Casc1* has 16 exons, spans a 36-kb region, and is flanked at its 3′-end by the 3′-end of *Lrmp* and at its 5′-end by the 5′-end of *Ghiso*.

A study at multiple-organ expression panels (Zhang 2003) found that *Casc1* was expressed at higher levels in lung, testis and kidney in both A/J and C57BL/6J strains. They studied the cellular localization of *Casc1* product in three

different cell lines and found a cytoplasmic distribution. In the same study, *Casc1* showed allelic-specific effects in vitro and in a tumor growth assay in nude mice, and was proposed as the candidate *Pas1* gene. *Casc1* protein is 67% identical and 81% similar to the derived human *CASC1* protein (similar to human hypothetical protein FLJ10921, 30.41 Mb). Analysis of the putative secondary structure of the *Casc1* encoded protein predicted a coiled-coil domain in its N-terminal region (amino acids 9-59) (http://www.ch.embnet.org/software/COILS_form.html). The coiled-coil structure is one of the principal oligomerization motifs in proteins (Burkhard et al., 2001). This similarity, with the *Lrmp* encoded protein, suggests that the two proteins may either interact with each other or with a common substrate.

***Ghiso*:** (also named 4930469P12Rik) maps to location 146.057600-146.063383, has 3 exons and its transcript length is 2,436 bps. This gene encodes a growth hormone-inducible soluble protein (AF412299) (of 86 residues) with no known protein domain therefore it was previously named *Ghiso*. *Ghiso*'s intragenic polymorphisms play a functional role in lung tumor susceptibility.

The upstream intragenic regions of both *Casc1* and *Ghiso* genes consist of a common 176-bp genomic fragment that, in light of its short length and the flanking of the 5'-ends of two genes, might share a bidirectional promoter activity.

KRAS: v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog maps at location 145.307569-145.341093, has 5 exons and its transcript length is 752 bps. This gene, a Kirsten ras oncogene homolog from the mammalian ras gene family, encodes a protein that is a member of the small GTPase

superfamily. A single amino acid substitution results in a transforming protein implicated in various malignancies, including lung adenocarcinoma, mucinous adenoma, ductal carcinoma of the pancreas and colorectal carcinoma. Alternative splicing leads to variants encoding two isoforms that differ in the C-terminal region. Somatic mutations of the *Kras2* gene play a causal role in lung tumorigenesis (Johnson et al., 2001) and have also been found in chronic inflammation conditions, such as chronic pancreatitis and ulcerative colitis (Yanagisawa et al., 1993; Lang et al., 1999). A transgenic mouse expressing the human c-H-ras gene showed enhanced susceptibility to autoimmune hepatitis (Tsunematsu et al., 1997). Thus, the *Kras2* gene may be a candidate for both lung tumor susceptibility and inflammatory response.

Genomic *Kras2* polymorphisms might modulate gene mutability and, in turn, numbers of tumors after chemical carcinogen treatment. The functional involvement of *Kras2* in lung tumorigenesis has been demonstrated for its wild-type (Zhang et al., 2001) and mutated alleles, which we found expressed at an almost perfect 1:1 ratio in lung tumors, in contrast with the reported loss of heterozygosity of the wild-type *Kras2* allele in mouse lung tumors (Zhang et al., 2001).

***Lmna-rs1*:** (also named 4933403M22Rik) maps to location 146.244904-146.286944, it has 9 exons and the transcript length is 1,514 bps.

Protein sequence comparison by BLAST showed that this gene encodes a protein (of 413 residues) containing an intermediate filament tail domain related to lamin A. For this reason the gene was previously named *Lmna-rs1* (lamin A-related sequence 1). It may be a tumor suppressor. Amino-

acid changes and down regulation are associated with lung adenoma susceptibility.

4.1 *Pas1* haplotype

The haplotype of the *Pas1* region defined strain relatedness in which the laboratory strains resolved as two groups according to the A/J- and the C57BL/6J-type haplotype, respectively.

This finding indicates that the *Pas1* QTL is due to an ancestral difference common to many laboratory strains rather than to a rare mutation specific to a single strain (Manenti et al., 1999).

Recent studies indicate that regions of low SNP density (~0.5 SNPs per 10 kb) are separated by regions of high density (~40 SNPs per 10 kb), creating a mosaic structure of genetic variations in mice and defining haplotype blocks of an average size of 1 Mb (Wade et al., 2002; Wiltshire et al., 2003). Laboratory strain haplotypes recapitulate population history of these strains and identify strain relatedness (Wiltshire et al., 2003). The *Pas1* core haplotype showed a high frequencies of polymorphisms (~100 SNPs/SSLPs per 10 kb) but resulted in only two haplotype-types. The *Pas1*-derived strain grouping was very different from that defined by whole-genome analysis of SNPs in mouse strains, which reflects strain genealogy (Wiltshire et al., 2003). Indeed, unlike the strain relatedness of the *Pas1* locus, whole-genome SNP analysis placed strains carrying the A/J- and C57BL/6J-type haplotype at the *Pas1* locus close together (see Fig. 4D of Wiltshire et al., 2003), where RF/J and AKR/J, CBA/J and C3H, SWR/J and SJL/J strains are placed in the same sub-branches. Therefore, the *Pas1* haplotype is not the product of strain genealogy.

In addition, 12/54 (22%) markers located within the haplotype did not cosegregate with the haplotype, suggesting that recombination had taken place during the creation of the haplotype block and that

linkage disequilibrium between the haplotype and non-essential polymorphisms has been lost, respectively, and two silent polymorphisms in the coding region of the *Lrmp* (*D6Int32c*) and *Kras2* gene (*D6Int142*).

Within the haplotype block, five amino-acid changes in three genes showed tight linkage with the haplotype (Tab. 8) and produced nonconservative changes.

The effect of any polymorphism on protein function relies not only on the type of change, but also on where the change occurs. For example, a change in a substrate binding site domain seems more likely to affect protein biochemical activity than a change in a non-critical domain. In any case, the prediction of a functional effect of any amino acid change awaits experimental evidence to distinguish the functional variations from the thousands of non functional genetic variations identified so far.

As the amino acid polymorphisms of three genes (*Lrmp*, *Casc1*, and *Lmna-rs1*), as well as intragenic polymorphisms of the *Bcat1*, *Ghiso*, and *Kras2* genes, are tightly linked in a single haplotype type, we cannot exclude any of these six genes as putative *Pas1* candidate genes. Moreover, 7 intergenic and 29 intragenic polymorphisms showed a tight linkage with the *Pas1* haplotype (Tab. 8).

We found that A/J and C57BL/6J alleles of recombinant *Lrmp* protein are distinguished by three amino acid variations, including L537P (*D6Int8*), which shows tight linkage not only with the *Pas1* locus haplotype, but also with susceptibility to an inflammatory response in a phenotypically selected mouse line. Because *Lrmp* encodes the Jaw protein, which is overexpressed in pre-T, pre-B, and mature B cells (Behrens *et al.*, 1994), this gene might have a role in the inflammatory response.

The G28D (*D6Int10*) polymorphism that we studied in *Ghiso* showed no significant linkage with *Pas1* locus activity or with inflammatory response.

These two genes showed no allelic differences in expression in normal mouse lungs or lung tumors, whereas *Casc1* showed preferential expression in lung tumors of the A/J-derived allele.

4.2 Candidate genes functional analysis

We decided to verify the hypothesis that the *Pas1* locus is comprised of a cluster of six cancer modifier genes in the *in vitro* study of these candidate genes.

Ghiso carries an amino-acid polymorphism that does not cosegregate with mouse lung tumor susceptibility, nevertheless we decided to include *Ghiso* in the *in vitro* study because the genomic structure of the 5' UTR region of *Casc1* and *Ghiso* are separated by 169 bp suggesting that they might share a unique bidirectional promoter and/or other regulatory elements.

By assaying the putative cancer modifier function of *Pas1* candidate genes in human lung cancer cell lines of different histotype origin and with different morphological and growth characteristics, we sought to provide clues into the possible effects of mouse genes in human cells as a step toward the application of findings in mice to humans.

Colony formation assay of cancer cell lines expressing allelic variants of candidate cancer modifier genes by transfection represents a convenient method of assessing functional in vitro activity of allelic variants, since the number of resulting colonies provides an estimate of the growth or survival of transfected cancer cells.

Results obtained by semi-quantitative Giemsa-stained colony formation were consistent with those obtained by cell count assay in both human lung cancer cell lines. The amino acid variants of *Lrmp* and *Ghiso* genes showed no significant differences in the number of

transfected colonies and cells, whereas *Casc1* and *Lmna-rs1* showed significant differences in allelic effects.

The observed lack of differential effects between the A/J and C57BL/6J alleles of the *Lrmp* and *Ghiso* genes on colony formation of transfected human cancer cell lines is consistent with previous observations in mouse cell lines obtained by Zhang *et al.* in 2003.

Our findings of a growth inhibitory effect of *Casc1* for A/J versus the C57BL/6J allele contrasts with findings by Zhang *et al.* in 2003, who instead found a stimulatory effect of the A/J *Casc1* in mouse cell lines. Since the recombinant transcripts used in both studies contain the same full-length coding regions of *Casc1*, there are no obvious explanations for the discrepant findings other than different somatic changes or difference species in the cell lines used. We obtained concordant results in two human lung cancer cell lines with a different KRAS mutation status: A549 cells carries a Gly→Ser mutation at the codon 12 of the KRAS gene (Valenzuela and Groffen, 1986), whereas NCI-H520 cells carries no KRAS mutation at codon 12 or 13 (data not shown).

We found allelic effects for the *Lmna-rs1* with a growth stimulatory effect of the A/J allele, whereas Zhang *et al.* in 2003 reported no allelic differences for the same full-length transcript that we assayed. However, the same authors recently reported the stimulatory effect of an A/J-specific mRNA isoform of *Lmna-rs1* in NIH mouse fibroblast cell lines (Whang *et al.*, 2005). *Lmna-rs1* may be a cell-type differentiation marker, whose expression is restricted to specific lung cell types that do not give rise to lung tumors.

We found that *Lmna-rs1* gene expression is not detectable in primary mouse lung tumors. These results suggest that *Lmna-rs1* is expressed in the early stages of lung tumorigenesis and becomes down regulated during tumor growth, so that the A/J allele may stimulate growth of initiated neoplastic cells, thus contributing to the effects of the *Pas1* locus.

In vitro assays cannot detect all the possible gene effects on tumorigenesis, which is a complex phenomenon involving tumor-host interactions modulated by numerous factors. After transfection of either the *Lmna-rs1* or *Casc1* allele we can exclude that inhibition of clonogenicity in vitro by allelic forms of *Pas1* candidate genes is mediated by apoptosis.

Further studies are needed to clarify the precise mechanisms of allele-specific in vitro effects of these two genes.

4.3 Inflammatory response

Lrmp, *Bcat1*, and *Kras2* have been associated with apoptosis, tumor development and inflammatory response (Eden & Benvenisty, 1999; Johnson et al., 2001; Maria et al., 2003).

In Maria 2003 it was shown that mouse outbred lines phenotypically selected for high (AIR_{max}) and low (AIR_{min}) response to s.c. nonspecific acute inflammation showed dramatic differences in their susceptibility to lung tumorigenesis. In particular AIR_{min} mice were ~40-fold more susceptible than AIR_{max} mice. It was found that in AIR lines there is a correlation between persistent subacute lung inflammation and susceptibility to lung tumorigenesis. LD analysis in AIR lines was performed using genetic markers located within the *Pas1* locus to assess the possible involvement of the *Pas1* locus in the inflammatory response. Markers located within a 452-kb *Pas1* region showed a highly significant LD pattern (-log P>8.0). Markers outside this region showed either nonsignificant or borderline LD values, indicating that the location of the gene(s) responsible for the AIR phenotype lies within this 452-kb region. The results of this work indicate that a genetic polymorphism(s) of at least one of the five genes is functionally responsible for the *Pas1* and AIR phenotypes.

4.4 Conclusion

The demonstration of allele-specific in vitro effects in human lung cancer cell lines of two of the four genes that carry amino acid variants and are included in the *Pas1* haplotype provides evidence that alleles of mouse candidate cancer modifier genes can modulate growth of human cancer cells. Overall, the in vitro effects of the A/J allele of the *Casc1* gene, unlike the in vivo effects of the *Pas1* haplotype in stimulating lung tumor development, and the absence of in vitro modulation activity of the *Lrmp* gene whose polymorphisms are tightly linked with the *Pas1* haplotype, support the possibility of a functional role for the whole *Pas1* haplotype, and not just of the individual genes comprising it, in determining genetic predisposition to lung cancer.

The results obtained in this work suggest that *Pas1* locus is comprised of a cluster of six cancer modifier genes, whose amino-acid and genomic polymorphisms operate to confer susceptibility or resistance to lung tumorigenesis.

Zhang *et al.* in 2003 proposed the *Casc1* gene as the *Pas1* gene. Moreover, in their study, genetic recombination did not separate the *Casc1* gene from the other five genes of the herein identified *Pas1* haplotype, confirming our hypothesis that none of the six genes can yet be formally excluded from *Pas1* candidacy.

4.5 Future aims

In conclusion, we achieved the aims of the project and now we continue to study the candidate genes.

First, we will study the putative bidirectional promoter that maps in the 169 bp region between *Casc1* and *Ghiso* by subcloning this region in pGL3 Luciferase Reporter Vector. pGL3 Vectors provide a basis for the quantitative analysis of factors that potentially regulate mammalian gene expression and contain a modified coding region for firefly (*Photinus pyralis*) luciferase that has been optimized for monitoring transcriptional activity in transfected eukaryotic cells.

Preliminary data, obtained in our laboratory, shown that *Casc1*, *Kras2* and *Lmna-rs1* genes give rise to a series of isoforms. Probably qualitative and quantitative variations of these isoforms might modulate the different susceptibility to lung carcinogenesis in mice. We will carried out an expression analysis of the different isoforms and then we will test their functional activity in vitro.

Our in vitro assays cannot detect all possible gene effects on tumorigenesis, which is a complex phenomenon involving tumor-host interactions modulated by numerous factors (Hanahan and Weinberg, 2000). Therefore, we will test tumor growth in vivo and then clone the human homologous genes, caring out an analysis of association between genetic polymorphism and lung adenocarcinoma risk and prognosis in a human population.
Then we will perform a detailed genetic analysis of the homologous human region in lung adenocarcinoma patients and healthy population controls. This might determine whether specific haplotypes also occur in humans and whether they control lung cancer risk.

Chapter 5 BIBLIOGRAPHY

ABECASIS GR, COOKSON WO. (2000). GOLD-graphical overview of linkage disequilibrium. Bioinformatics.; **16(2)**: 182-3.

ANDERVONT HB, SHIMKIN MB. (1940). J. Natl. Cancer Inst.; **1**: 225-239.

AKAMATSU Y. (1975). Neoplasms in strains of splenectomized mice after a single 7,12-dimethylbenz[alpha]anthracene treatment. J Natl Cancer Inst.; **55(4)**: 893-7.

ASKLING J, GRUNEWALD J, EKLUND A, HILLERDAL G, EKBOM A. (1999). Increased risk for cancer following sarcoidosis. Am J Respir Crit Care Med.; **160(5 Pt 1)**: 1668-72.

BALE AE, LI FP (1997). Priciples of cancer management: Cancer genetics. Cancer: principles and practice of oncology. Fifth edition.; **14**: 285-293.

BALMAIN A, NAGASE H. (1998).Cancer resistance genes in mice: models for the study of tumour modifiers. Trends Genet.; **14(4)**: 139-44.

BALMAIN A. (2002). Cancer as a complex genetic trait: tumor susceptibility in humans and mouse models. Cell.; **108(2)**: 145-52.

BAUER AK, DWYER-NIELD LD, KEIL K, KOSKI K, MALKINSON AM. (2001). Butylated hydroxytoluene (BHT) induction of pulmonary inflammation: a role in tumor promotion. Exp Lung Res.; **27(3)**: 197-216.

BEHRENS TW, JAGADEESH J, SCHERLE P, KEARNS G, YEWDELL J,

STAUDT LM. (1994). Jaw1, A lymphoid-restricted membrane protein localized to the endoplasmic reticulum. J Immunol.; **153(2)**:682-90.

BENTVELZEN PAJ. AND SZALAYN G. (1966). Some genetic aspects of differences in susceptibility to the development of lung tumours between inbred strains of mice. In Severi,L. (ed.) Lung Tumours in Animals; University of Perugia, Perugia, Italy: pp. 835–844.

BOTSTEIN D, RISCH N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet.; **33 Suppl**: 228-37.

BROWNSON RC, ALAVANJA MC. (2000). Previous lung disease and lung cancer risk among women (United States). Cancer Causes Control.; **11(9)**: 853-8.

CAREY JC, LAUB JM, HALL BD. (1979). Penetrance and variability in neurofibromatosis: a genetic study of 60 families. Birth Defects Orig Artic Ser.; **15(5B)**: 271-81.

COLLINS FS, GUYER MS, CHARKRAVARTI A. (1997). Variations on a theme: cataloging human DNA sequence variation. Science.; **278(5343)**: 1580-1.

CONNER BJ, REYES AA, MORIN C, ITAKURA K, TEPLITZ RL, WALLACE RB. (1983). Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides. Proc Natl Acad Sci U S A.; **80(1)**: 278-82.

DARVASI A, PISANTE-SHALOM A. (2002). Complexities in the

genetic dissection of quantitative trait loci. Trends Genet.; **18(10)**: 489-91.

DEVEREUX TR, WISEMAN RW, KAPLAN N, GARREN S, FOLEY JF, WHITE CM, ANNA C, WATSON MA, PATEL A, JARCHOW S, ET AL. (1994). Assignment of a locus for mouse lung tumor susceptibility to proximal chromosome 19. Mamm Genome.; **5(12)**: 749-55.

DIETRICH W, KATZ H, LINCOLN SE, SHIN HS, FRIEDMAN J, DRACOPOLI NC, LANDER ES. (1992). A genetic map of the mouse suitable for typing intraspecific crosses. Genetics.; **131(2)**: 423-47.

DRAGANI TA, CANZIAN F, PIEROTTI MA. (1996). A polygenic model of inherited predisposition to cancer. FASEB J.;**10(8)**: 865-70.

DRAGANI TA, MANENTI G, PIEROTTI MA. (1995). Genetics of murine lung tumors. Adv Cancer Res.; **67** :83-112.

DRAGANI TA. (2003). 10 years of mouse cancer modifier loci: human relevance. Cancer Res.; **63(12)**: 3011-8.

EASTON DF, PONDER MA, HUSON SM, PONDER BA. (1993). An analysis of variation in expression of neurofibromatosis (NF) type 1 (NF1): evidence for modifying genes. Am J Hum Genet.; **53(2)**: 305-13.

EDEN A, BENVENISTY N. (1999). Involvement of branched-chain amino acid aminotransferase (Bcat1/Eca39) in apoptosis. FEBS Lett.; **457(2)**: 255-61.

EMERY J, LUCASSEN A, MURPHY M. (2001). Common hereditary cancers and implications for primary care. Lancet.; **358(9275)**: 56-63.

FALCONER DS. AND BLOOM JL. (1962). A genetic study of induced lung tumours in mice. Br. J. Cancer; **16**: 665–6

FESTING MF, YANG A, MALKINSON AM. (1994). At least four genes and sex are associated with susceptibility to urethane-induced pulmonary adenomas in mice. Genet Res.; **64(2)**: 99-106.

FESTING MF. AND BLACKMORE DK. (1971). Life span of specified-pathogen-free (MRC category 4) mice and rats. Lab. Anim.; **5**: 179–192.

FONG KM, SEKIDO Y, GAZDAR AF, MINNA JD. (2005). Lung cancer. 9: Molecular biology of lung cancer: clinical implications. Thorax.; **58(10)**: 892-900.

GALBIATI F, PETTINICCHIO A, DRAGANI TA, MANENTI G. (2006). Allelic effects of mouse *Pas1* candidate genes in human lung cancer cell lines. Cancer Lett.; 244(2):176-81.

GARIBOLDI M, MANENTI G, CANZIAN F, FALVELLA FS, RADICE MT, PIEROTTI MA, DELLA PORTA G, BINELLI G, DRAGANI TA. (1993). A major susceptibility locus to murine lung carcinogenesis maps on chromosome 6. Nat Genet.; **3(2)**: 132-6.

GAZDAR AF, MINNA JD. (1999). Molecular detection of early lung cancer. J Natl Cancer Inst.; **91(4)**: 299-301.

HANAHAN D, WEINBERG RA. (2000). The hallmarks of cancer. Cell.;

100(1): 57-70.

HEWETT D, SAMUELSSON L, POLDING J, ENLUND F, SMART D, CANTONE K, SEE CG, CHADHA S, INEROT A, ENERBACK C, MONTGOMERY D, CHRISTODOLOU C, ROBINSON P, MATTHEWS P, PLUMPTON M, WAHLSTROM J, SWANBECK G, MARTINSSON T, ROSES A, RILEY J, PURVIS I. (2002). Identification of a psoriasis susceptibility candidate gene by linkage disequilibrium mapping with a localized single nucleotide polymorphism map. Genomics.; **79(3)**: 305-14.

HIRSCH FR, FRANKLIN WA, GAZDAR AF, BUNN PA JR. (2001). Early detection of lung cancer: clinical perspectives of recent advances in biology and radiology. Clin Cancer Res.; **7(1)**: 5-22.

HIRST GL, BALMAIN A. (2004). Forty years of cancer modelling in the mouse. Eur J Cancer.; **40(13)**: 1974-80.

HOAG WG. (1963). Spontaneous cancer in mice. Ann N Y Acad Sci.; **108**: 805-31.

HORIKAWA Y, ODA N, COX NJ, LI X, ORHO-MELANDER M, HARA M, HINOKIO Y, LINDNER TH, MASHIMA H, SCHWARZ PE, DEL BOSQUE-PLATA L, HORIKAWA Y, ODA Y, YOSHIUCHI I, COLILLA S, POLONSKY KS, WEI S, CONCANNON P, IWASAKI N, SCHULZE J, BAIER LJ, BOGARDUS C, GROOP L, BOERWINKLE E, HANIS CL, BELL GI. (2000). Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet.; **26(2)**: 163-75.

JOHNSON L, MERCER K, GREENBAUM D, BRONSON RT, CROWLEY D, TUVESON DA, JACKS T. (2001). Somatic activation of the K-

ras oncogene causes early onset lung cancer in mice. Nature.; **410(6832)**: 1111-6.

KIMURA T, YOSHIDA K, SHIMADA A, JINDO T, SAKAIZUMI M, MITANI H, NARUSE K, TAKEDA H, INOKO H, TAMIYA G, SHINYA M. (2005). Genetic linkage map of medaka with polymerase chain reaction length polymorphisms. Gene.; **363**: 24-31.

KIRK BW, FEINSOD M, FAVIS R, KLIMAN RM, BARANY F. (2002). Single nucleotide polymorphism seeking long term association with complex disease. Nucleic Acids Res.; **30(15)**: 3295-311.

KLAUSNER RD. (1999). Studying cancer in the mouse. Oncogene; **18(38)**: 5249-52.

KORSTANJE R, PAIGEN B. (2002). From QTL to gene: the harvest begins. Nat Genet.; **31(3)**: 235-6.

LAI E, RILEY J, PURVIS I, ROSES A. (1998). A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. Genomics.; **54(1)**: 31-8.

LANDER ES, SCHORK NJ. (1994). Genetic dissection of complex traits. Science.; **265(5181)**: 2037-48.

LANDER ES. (1996). The new genomics: global views of biology. Science.; **274(5287)**: 536-9.

LANG SM, STRATAKIS DF, HEINZLMANN M, HELDWEIN W, WIEBECKE B, LOESCHKE K. (1999). Molecular screening of patients with long standing extensive ulcerative colitis: detection of p53 and Ki-ras mutations by single strand

conformation polymorphism analysis and differential hybridisation in colonic lavage fluid. Gut.; **44(6)**: 822-5.

LYNCH CI (1926). J Exp. Med.; **43**: 339-335

LYNCH M, WALSH B. (1997). Genetic and analysis of quantitative traits. Sinauer Associates, Sunderland, MA

MAHLER JF., STOKES W., MANN PC., TAKAOKA M. AND MARONPOT RR. (1996). Spontaneous lesions in aging FVB/N mice. Toxicol. Pathol.; **24**: 710–716

MALKINSON AM, BEER DS. (1983). Major effect on susceptibility to urethan-induced pulmonary adenoma by a single gene in BALB/cBy mice. J Natl Cancer Inst.; **70(5)**: 931-6.

MALKINSON AM, BEER DS. (1984). Pharmacologic and genetic studies on the modulatory effects of butylated hydroxytoluene on mouse lung adenoma formation. J Natl Cancer Inst.; **73(4)**: 925-33.

MALKINSON AM. (1989). The genetic basis of susceptibility to lung tumors in mice. Toxicology.; **54(3)**: 241-71.

MALTONI C., CILIBERTI A., COTTI G., CONTI B. AND BELPOGGI F. (1989). Benzene, an experimental multipotential carcinogen: results of the long-term bioassays performed at the Bologna Institute of Oncology. Environ. Health Perspect.; **82**: 109–124.

MANENTI G, ACEVEDO A, GALBIATI F, GIANNI BARRERA R, NOCI S, SALIDO E, DRAGANI TA. (2002). Cancer modifier alleles inhibiting lung tumorigenesis are common in mouse inbred

strains. Int J Cancer.; **99(4)**: 555-9.

MANENTI G, FALVELLA FS, GARIBOLDI M, DRAGANI TA, PIEROTTI MA. (1995) Different susceptibility to lung tumorigenesis in mice with an identical Kras2 intron 2. Genomics.; **29(2)**: 438-44.

MANENTI G, GALBIATI F, GIANNI-BARRERA R, PETTINICCHIO A, ACEVEDO A, DRAGANI TA. (2004) Haplotype sharing suggests that a genomic segment containing six genes accounts for the pulmonary adenoma susceptibility 1 (*Pas1*) locus activity in mice. Oncogene.; **23(25)**: 4495-504.

MANENTI G, GARIBOLDI M, ELANGO R, FIORINO A, DE GREGORIO L, FALVELLA FS, HUNTER K, HOUSMAN D, PIEROTTI MA, DRAGANI TA. (1996). Genetic mapping of a pulmonary adenoma resistance (Par1) in mouse. Nat Genet.; **12(4)**: 455-7.

MANENTI G, STAFFORD A, DE GREGORIO L, GARIBOLDI M, FALVELLA FS, AVNER P, DRAGANI TA. (1999). Linkage disequilibrium and physical mapping of *Pas1* in mice. Genome Res.; **9(7)**: 639-46.

MANENTI G., GARIBOLDI M., FIORINO A., ZEDDA AI., PIEROTTI M.A. AND DRAGANI T.A. (1997). *Pas1* is a common lung cancer susceptibility locus in three mouse strains. Mamm. Genome; **8**: 801–809.

MARIA DA, MANENTI G, GALBIATI F, RIBEIRO OG, CABRERA WH, BARRERA RG, PETTINICCHIO A, DE FRANCO M, STAROBINAS N, SIQUEIRA M, DRAGANI TA, IBANEZ OM.

(2003). Pulmonary adenoma susceptibility 1 (*Pas1*) locus affects inflammatory response. Oncogene.; **22(3)**: 426-32.

MAYNE ST, BUENCONSEJO J, JANERICH DT. (1999). Familial cancer history and lung cancer risk in United States nonsmoking men and women. Cancer Epidemiol Biomarkers Prev.; **8(12)**: 1065-9.

MINNA JD, ROTH JA, GAZDAR AF. (2002). Focus on lung cancer. Cancer Cell.; **1(1)**:49-52.

MULLER-MYHSOK B, ABEL L. (1997). Genetic analysis of complex diseases. Science.; **275(5304)**: 1328-9; author reply 1329-30.

OOI WL, ELSTON RC, CHEN VW, BAILEY-WILSON JE, ROTHSCHILD H. (1986). Increased familial risk for lung cancer. J Natl Cancer Inst.; **76(2)**: 217-22.

PARK IW, WISTUBA II, MAITRA A, MILCHGRUB S, VIRMANI AK, MINNA JD, GAZDAR AF. (1999). Multiple clonal abnormalities in the bronchial epithelium of patients with lung cancer. J Natl Cancer Inst.; **91(21)**: 1863-8.

PETO R, CHEN ZM, BOREHAM J. (1999). Tobacco--the growing epidemic. Nat Med.; **5(1)**: 15-7.

PONDER BA (2001). Cancer genetics. Nature.; **411(6835)**: 336-41.

PONDER BA. (1990). Inherited predisposition to cancer. Trends Genet.; **6(7)**: 213-8.

RIOUX JD, DALY MJ, SILVERBERG MS, LINDBLAD K, STEINHART H, COHEN Z, DELMONTE T, KOCHER K, MILLER K, GUSCHWAN

S, KULBOKAS EJ, O'LEARY S, WINCHESTER E, DEWAR K, GREEN T, STONE V, CHOW C, COHEN A, LANGELIER D, LAPOINTE G, GAUDET D, FAITH J, BRANCO N, BULL SB, MCLEOD RS, GRIFFITHS AM, BITTON A, GREENBERG GR, LANDER ES, SIMINOVITCH KA, HUDSON TJ. (2001). Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet.; **29(2)**: 223-8.

RISCH N, MERIKANGAS K. (1996). The future of genetic studies of complex human diseases. Science.; **273(5281)**:1516-7.

RIVERA MP. (2004). Multimodality therapy in the treatment of lung cancer. Semin Respir Crit Care Med.; **25 Suppl** 1:3-10.

ROSES AD. (2000). Pharmacogenetics and pharmacogenomics in the discovery and development of medicines. Novartis Found Symp. **229**: 63-6; discussion 66-70.

RUSSO A, ZANNA I, TUBIOLO C, MIGLIAVACCA M, BAZAN V, LATTERI MA, TOMASINO RM, GEBBIA N. (2000). Hereditary common cancers: molecular and clinical genetics. Anticancer Res.; **20(6C)**: 4841-51.

RYAN J, BARKER PE, NESBITT MN, RUDDLE FH. (1987). KRAS2 as a genetic marker for lung tumor susceptibility in inbred mice. J Natl Cancer Inst.; **79(6)**: 1351-7.

SHAFARENKO M., MAHLER J., COCHRAN C., KISIELEWSKI A., GOLDING E., WISEMAN R. AND GOODROW T. (1997). Similar incidence of K-ras mutations in lung carcinomas of FVB/N mice and FVB/N mice carrying a mutant p53 transgene. Carcinogenesis; **18**: 1423–1426

SHAUGHNESSY JD JR, LARGAESPADA DA, TIAN E, FLETCHER CF, CHO BC, VYAS P, JENKINS NA, COPELAND NG. (1999). Mrvi1, a common MRV integration site in BXH2 myeloid leukemias, encodes a protein with homology to a lymphoid-restricted membrane protein Jaw1. Oncogene.; **18(12)**: 2069-84.

SHIMKIN MB. (1940). J. Natl. Cancer Inst.; **1**: 241-254

SHIMKIN MB, STONER GD. (1975). Lung tumors in mice: application to carcinogenesis bioassay. Adv Cancer Res.; **21**: 1-58.

SHIMKIN MB. (1955). Pulmonary tumors in experimental animals. Adv Cancer Res.; **3**: 223-67.

SMITH GS., WALFORD RL. AND MICKEY MR. (1973). Lifespan and incidence of cancer and other diseases in selected long-lived inbred mice and their F_1 hybrids. J. Natl Cancer Inst.; **50**: 1195–1213.

SNYDER HL, BACIK I, BENNINK JR, KEARNS G, BEHRENS TW, BACHI T, ORLOWSKI M, YEWDELL JW. (1997). Two novel routes of transporter associated with antigen processing (TAP)-independent major histocompatibility complex class I antigen processing. J Exp Med.; **186(7)**: 1087-98.

THAETE LG, MALKINSON AM. (1991). Cells of origin of primary pulmonary neoplasms in mice: morphologic and histochemical studies. Exp Lung Res.; **17(2)**: 219-28.

THAETE LG., NESBITTN MN. AND MALKINSON AM. (1991). Lung adenoma structure among inbred strains of mice: the pulmonary adenoma histologic type (Pah) genes. Cancer Lett.; **61**: 15–20.

THOMAS P, SMART TG. (2005). HEK293 cell line: a vehicle for the expression of recombinant proteins. J Pharmacol Toxicol Methods.; **51(3)**: 187-200.

TOKUHATA GK, LILIENFELD AM. (1963). Familial aggregation of lung cancer among hospital patients. Public Health Rep.; **78**: 277-83.

TOKUHATA GK, LILIENFELD AM. (1963). Familial aggregation of lung cancer in humans. J Natl Cancer Inst.; **30**: 289-312.

TRAININ N, LINKER-ISRAELI M. (1967). Restoration of immunologic reactivity of thymectomized mice by calf thymus extracts. Cancer Res.; **27(2)**: 309-13.

TRAININ N., PRECERUTTI A. AND LAW LW. (1964). Trends in carcinogenesis by urethan administration to new-born mice of different strains. Nature; **202**: 305–306.

TRIPODIS N, HART AA, FIJNEMAN RJ, DEMANT P. (2001). Complexity of lung cancer modifiers: mapping of thirty genes and twenty-five interactions in half of the mouse genome. J Natl Cancer Inst.; **93(19)**: 1484-91.

TSUNEMATSU S, SAITO H, TADA S, EBINUMA H, TSUCHIYA M, KUMAGAI N, MORIZANE T, NOMURA T, ISHII H. (1997). Susceptibility of experimental autoimmune hepatitis in transgenic mice overexpressing the c-H-ras gene. J Gastroenterol Hepatol.; **12(4)**: 319-24.

TU IP, WHITTEMORE AS. (1999). Power of association and linkage tests when the disease alleles are unobserved. Am J Hum

Genet.; **64(2)**: 641-9.

TUVESON DA, JACKS T. (1999). Modeling human lung cancer in mice: similarities and shortcomings. Oncogene.; **18(38)**: 5318-24.

VALENZUELA DM, GROFFEN J. (1986). Four human carcinoma cell lines with novel mutations in position 12 of c-K-ras oncogene. Nucleic Acids Res.; **14(2)**: 843-52.

VAN DER VALK MA. (1981). Survival, tumor incidence and gross pathology in 33 mouse strains. In Hilgers,J. and Sluyser,M. (eds) Mammary Tumors in the Mouse.; Elsevier, Amsterdam, The Netherlands, pp. 46–115.

WADE CM, KULBOKAS EJ 3RD, KIRBY AW, ZODY MC, MULLIKIN JC, LANDER ES, LINDBLAD-TOH K, DALY MJ. (2002). The mosaic structure of variation in the laboratory mouse genome. Nature.; **420(6915)**: 574-8.

WANG M, FUTAMURA M, WANG Y, YOU M. (2005). *Pas1*c1 is a candidate for the mouse pulmonary adenoma susceptibility 1 locus. Oncogene.; **24(11)**: 1958-63.

WANG M, LEMON WJ, LIU G, WANG Y, IRAQI FA, MALKINSON AM, YOU M. (2003). Fine mapping and identification of candidate pulmonary adenoma susceptibility 1 genes using advanced intercross lines. Cancer Res.; **63(12)**: 3317-24.

WEISSMAN SM. (1995). Genetic bases for common polygenic diseases. Proc Natl Acad Sci U S A.; **92(19)**: 8543-4.

WILTSHIRE T, PLETCHER MT, BATALOV S, BARNES SW, TARANTINO

LM, COOKE MP, WU H, SMYLIE K, SANTROSYAN A, COPELAND NG, JENKINS NA, KALUSH F, MURAL RJ, GLYNNE RJ, KAY SA, ADAMS MD, FLETCHER CF. (2003). Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. Proc Natl Acad Sci U S A.; **100(6)**: 3380-5.

YANAGISAWA A, OHTAKE K, OHASHI K, HORI M, KITAGAWA T, SUGANO H, KATO Y. (1993). Frequent c-Ki-ras oncogene activation in mucous cell hyperplasias of pancreas suffering from chronic inflammation. Cancer Res.; **53(5)**: 953-6.

ZHANG Z, FUTAMURA M, VIKIS HG, WANG M, LI J, WANG Y, GUAN KL, YOU M. (2003). Positional cloning of the major quantitative trait locus underlying lung tumor susceptibility in mice. Proc Natl Acad Sci U S A.; **100(22)**: 12642-7.

ZHANG Z, WANG Y, VIKIS HG, JOHNSON L, LIU G, LI J, ANDERSON MW, SILLS RC, HONG HL, DEVEREUX TR, JACKS T, GUAN KL, YOU M (2001). Wildtype Kras2 can inhibit lung carcinogenesis in mice. Nat Genet.; **29(1)**: 25-33.

NOTE: MATERIAL PUBLISHED FROM THIS THESIS

Published from Chapter 3, Section 3.1: *Pas1* CONTAINING REGION and Section 3.2: CANDIDATE GENES FUNCTIONAL ANALISYS:

MARIA DA, MANENTI G, **GALBIATI F**, RIBEIRO OG, CABRERA WH, BARRERA RG, PETTINICCHIO A, DE FRANCO M, STAROBINAS N, SIQUEIRA M, DRAGANI TA, IBANEZ OM. (2003). Pulmonary adenoma susceptibility 1 (*Pas1*) locus affects inflammatory response. Oncogene.; **22(3)**: 426-32.

MANENTI G, **GALBIATI F**, GIANNI-BARRERA R, PETTINICCHIO A, ACEVEDO A, DRAGANI TA. (2004) Haplotype sharing suggests that a genomic segment containing six genes accounts for the pulmonary adenoma susceptibility 1 (*Pas1*) locus activity in mice. Oncogene.; **23(25)**: 4495-504.

GALBIATI F, PETTINICCHIO A, DRAGANI TA, MANENTI G. (2006). Allelic effects of mouse *Pas1* candidate genes in human lung cancer cell lines. Cancer Lett.; **244(2)**:176-81.